(12) **UK Patent Application** (19) **GB** (11) **2 310 745** (13) **A**

(43) Date of A Publication 03.09.1997

(54) **Dynamic random access memory**

(57) A dynamic random access memory (DRAM) has short line segments 104, each of which is coupled to a separate block of memory cells 106 and a corresponding amplifier 108, which is activated in read mode, write mode and refresh mode. The bit line segments are separated by pass transistors 110, which are enabled only in connection with data input and output. Very small cell capacitors can be used, allowing the use of conventional gate capacitors, without increasing power consumption. The speed of read and write operations may be very high. Further aspects relate to other DRAM circuits (Figs 2A-3J, 6A, 8A-H, 15, 16A-D, 19A-D), including read/write, refresh and clamping features, and to methods of forming a capacitor and/or a submerged transistor on a semiconductor wafer (Figs 9A-11C).

FIG. 1A.

GB 2 310 745 A

At least one drawing originally filed was informal and the print reproduced here is taken from a later filed formal copy.

FIG. IA.

TO WL
DECODERS-
-DRIVERS

TO
SEGMENTS

AMPLIFIER
CONTROL
SIGNALS

RAMP

WL ENABLE

SEGM. ENABLE

I/O MODE

144

STOP

REFRESH
COUNTER

142

140

COUNT

MEMORY SEQUENCE
GENERATOR AND
LOGIC

LOW ORDER
ADDRESS

112

SEGMENT ADDRESS

112

MEMORY REQUEST

MEMORY BUSY

READ/ WRITE

COPLIO

BLOC ADDR. ENABLE

IO LATCH CLAMP

THREE STATE CONTROL

BLOC ADDRESS

DATA I/O BIT I, 2 OR 4 BYTES

FIG. IB.

FIG. 2A.

FIG. 2B.

FIG. 2C.

VCLAMP

VDUM

VDDA

VSSA

VBLR

VBLW

VWL

VCELL
ZERO STORED

VCELL ONE STORED

VBLW

VBLR

VWL

FIG. 2D.

FIG. 2E.

BLR BLW

VCLAMP

302

304

VCLAMPL

FIG. 3A.

306

VCLAMP

304 302

VCLAMPL

FIG. 3B.

VCLAMP2

306

VCLAMPI

304 302

VCLAMPL

FIG. 3C.

BLR BLW

VCLAMP

306

VCLAMPB

308

VCLAMPL

FIG. 3D.

BLR   BLW

VCLAMP

306

VCLAMPB

310   312

VCLAMPHOLD

308

VCLAMPL

FIG. 3E.

VCLAMP2

306

VCLAMPI

304   302

VCLAMPL

VCLAMPB

310   312

FIG. 3F.

VCLAMP2

306

VCLAMPI

304   302

VCLAMPL

VCLODD

314   316

VCLEVEN

FIG. 3G.

FIG. 3H.

FIG. 3I.

FIG. 3J.

FIG. 4A.

——— TWO INVERTER TIMING
- - - SELECTED SEGMENT

VINV2
VCLAMP
BLR
BLW
WL
VCOUPL

- - - NON SELECTED SEGMENT

VCOUPL
VCLAMP
VINV2

FIG. 4B.

VCLAMP

412

422

VCLAMPB

VDD

406

402

VINV2

410

424

VINV2B

404   408

VSS

420   VCOUPL

NEXT
SEGMENT

FIG. 4C.

FIG. 5A.

SELINVTIME

NONSELINVTIME

VINV2

WORD LINE ENABLE

VRAMP

VDUM CKT → VDUM

SEGMENT
DECODER

VCLAMP
VCLAMPB

WRITE

REFRESH

WLONTIME

DUMMY ONTIME

EARLY CLAMP TIME

LATE CLAMP TIME

FIG. 5B.

FIG. 6A.

VCLAMPB

VCOUPL

VCLAMP

VCLAMPB

VCOUPL

BLW

BLR

WL

NOT ON BOTTOM
SEGMENT WHEN
WRITING AND
NOT SELECTED

VCOUPL

VCLAMP

VCLAMPB

VCLAMP
VCLAMPB

VCOUPL

BLR

BLW

FIG. 6B.

118

SEGMENT
DECODER

ODD/EVEN
SEGMENT

REFRESH B

FIG. 6C.

FIG. 7A.

VSTDB —

VDDA ODD

—

VSSA ODD

VCLAMP

VCOMP ODD

VWL ODD

VDDA EVEN

VSSA EVEN

VCOUPL

BL2 (REFRESH)

BLI

BL2

BL1 LAST SEGMENT

BL2

BLI

BL2 LAST SEGMENT

BLI (REFRESH)

FIG. 7B.

FIG. 7C.

SELECTED SEGMENT

VC

VINVODD
VBL 2
VCOUPL
VWLEVEN
VCOMPEVEN
VBL 1
VWLEVEN
VINVODD

NON SELECTED SEGMENT

VINVODD
VBL
VCLAMP
VCOMP EVEN
VBL 2
VCLAMP
VBL
VINVODD
VBL 1
VCOUPL

FIG. 7D.

FIG. 7E.

FIG. 8A.

FIG. 8B.

FIG. 8C.

FIG. 8D.

840

FIG. 8E.

VBIAS

VWLT

VBLT

READIN ONE   READ ONE   READ ZERO

DATA
OUT
ENABLE

DATA OUT

FIG. 8F.

FIG. 8G.

FIG. 8H.

I.

904 906 902

SUPPORT WAFER

2.

908

POLY THIN OXIDE

3.

4.

910

5.

912 FUSED AND THINNED WAFER

6.

916 914 916 914

OXIDE

918

7.

920 920 920

8.

922 GATE 922

FIG. 9A.

FIG. 9B.

SECTION B-B

N₁

N₂ SOURCE
TO N₁
DRAIN

B ⟶   ⟶ A

BLR

N₂   N₃

BLW

C   C

B ⟶   ⟶ A

SECTION A-A

N₃ TO
CAPACITOR

POLY CAP.
GROUND DIFFUSION

SECTION C-C

N+ DIFFUSIONS

SUPPORT WAFER

FIG. 9D.

POLY
CONNECTION   GATE   GATE POLY COVER

)THIN TRANSISTOR LAYER

POLY CAPACITOR

SUBSTRATE
CONNECTION

936

SUPPORT WAFER

FIG. 9C.

BLR₁   BLW₁   BLR₂ BLW₂

WLN+1

BLR
1

BLR
2

942

944

WLN

BLW
1

BLW
2

WLN-1

FIG. 9E.

SOURCE AND DRAIN
CONNECTION TO BUSSE

GROUND PLAIN (POLY)

CAPACITOR PLAIN CONNECTED
TO CONTROL NODE

N+ SOURCE/DRAIN

N+ SOURCE/DRAIN

SILICIDE

SILICIDE

CAPACITOR DIELECTRIC
(OXINITRIDE)

GATE CONNECTED TO
CONTROL NODE

FIG. 9F.

A) RECESS ETCH

1002

1004

B) N+ IMPLANT

1006

C) OXIDE GROUT AND POLY DEPOSITOR

1010

POLY

1008

D) AFTER POLISHING

1012
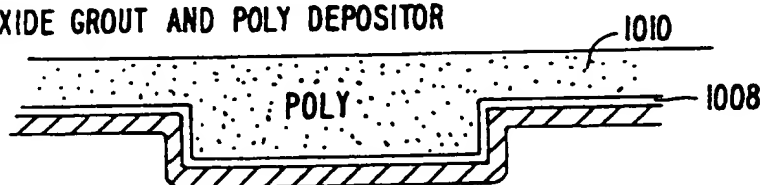
E) GROUND CONTACT AREA IMPLANT

1006

F) EPITAXIAL LAYER ADDITION

1014

}EPI LAYER

G) OXYGENE IMPLANT

EPI LAYER

}FOR SiO2

SiO2 ——
LAYER

DEVICES
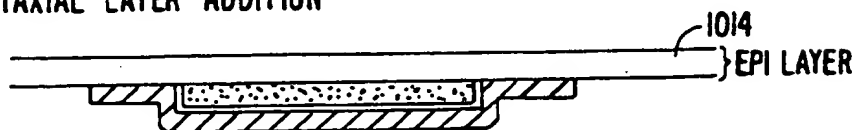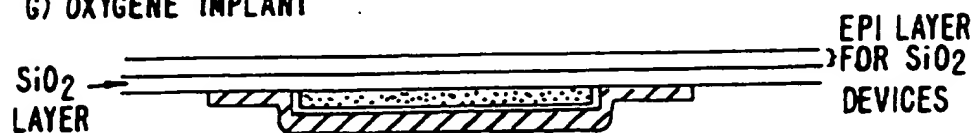
FIG. 10A.

A) AFTER RECESS ETCH THIN OXIDE AND POLY DEPOSITION
(IN LOCO OR BLANKET AND POLISH)

1016

B) SOURCE AND DRAIN IMPLANT

1018

C) EPI LAYER GROWTH

1020
} EPI LAYER

D) AFTER OXYGENE IMPLANT

} DEVICE LAYER
SiO$_2$

DRAIN
GATE AND
CAPACITOR
SOURCE
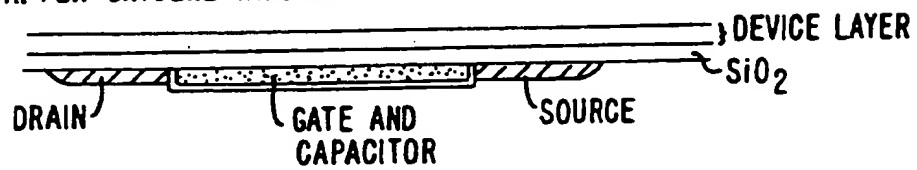
FIG. IOB.

A) WAFER WITH DEEP N+ DIFFUSION

1022
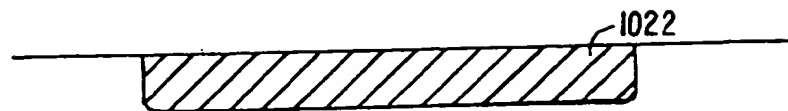
B) NITROGEN IMPLANT

1024
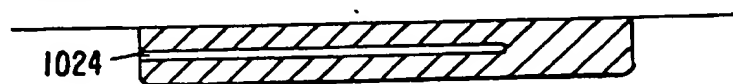
C) ISOLATION CUTS AND REFILL (SiO$_2$)

1026   1026

SiO$_2$

Si$_3$N$_4$ (SiO$_2$)

D) EPI LAYER GROWTH

1028   GROWN EPI (P–)

E) NITROGEN OR O$_2$ IMPLANT
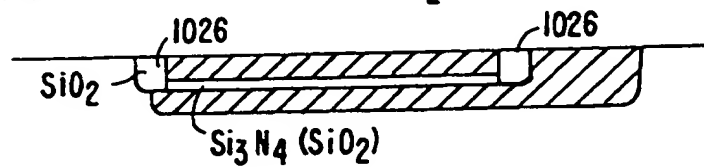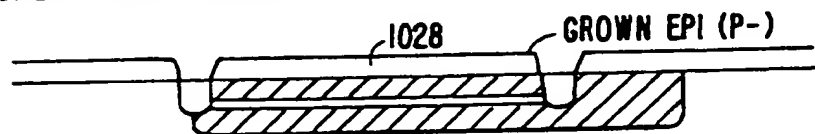
CAPACITOR
CONTACT   DEVICE LAYER   GROUND CONTACT
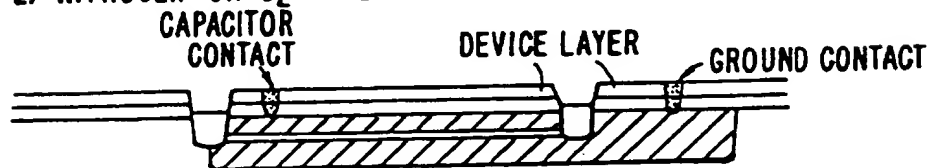
FIG. 10C.

A) TRANSITION DRAIN AND GROUND DIFFUSION (N+)
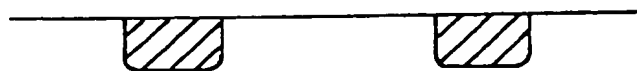
B) Si_3N_4 OR SiO_2 IMPLANT
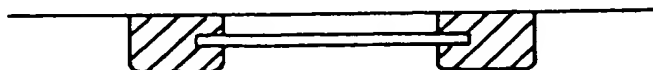
C) ISOLATION CUT AND REFILL
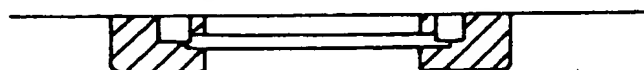
D) CAPACITOR N+ IMPLANT

E) P- EPI GROWTH

F) N_2 OR O_2 IMPLANT

DRAIN CONTACT   GATE CONTACT   SOURCE CONTACT (GND)

} DEVICE LAYER
THICK INSULATION

THIN INSULATION (GATE OXIDE)

FIG. 10D.

SUBMERGED N+ GROUND
(N₁ SOURCE)

BLR   BLW

wL

N2   N3   SUBMERGED GATE (CAP.)

N₂ SOURCE TO N₁
DRAIN CONNECTION

N3 SOURCE TO
NI GATE CONNECTION

SUBMERGED
N₁ DRAIN

wL

BLR   BLW   METAL BIT LINE TO N₃
DRAIN CONNECTION

FIG. IOE.
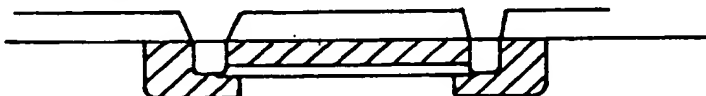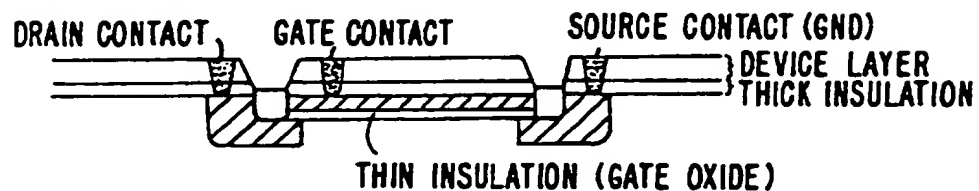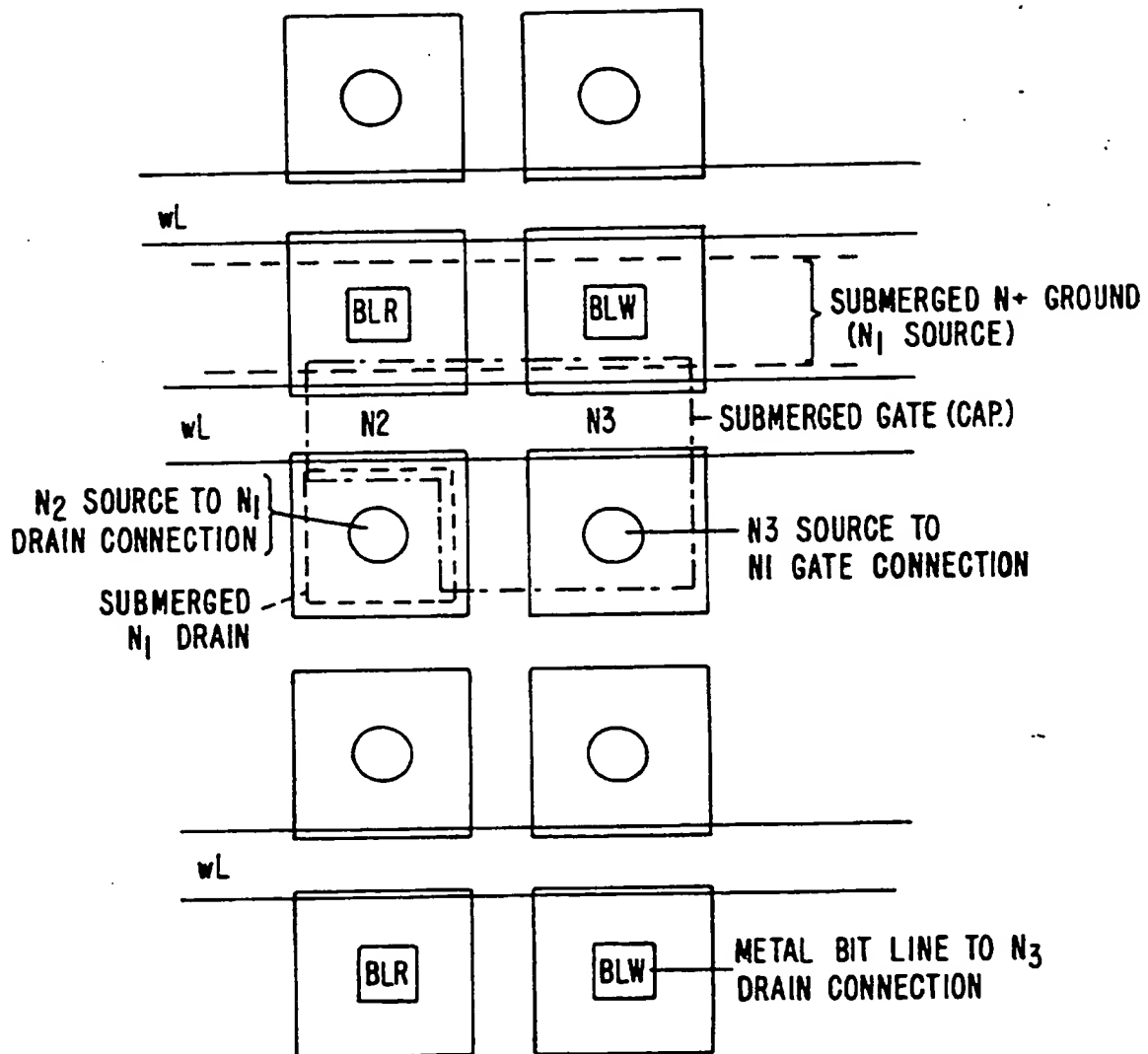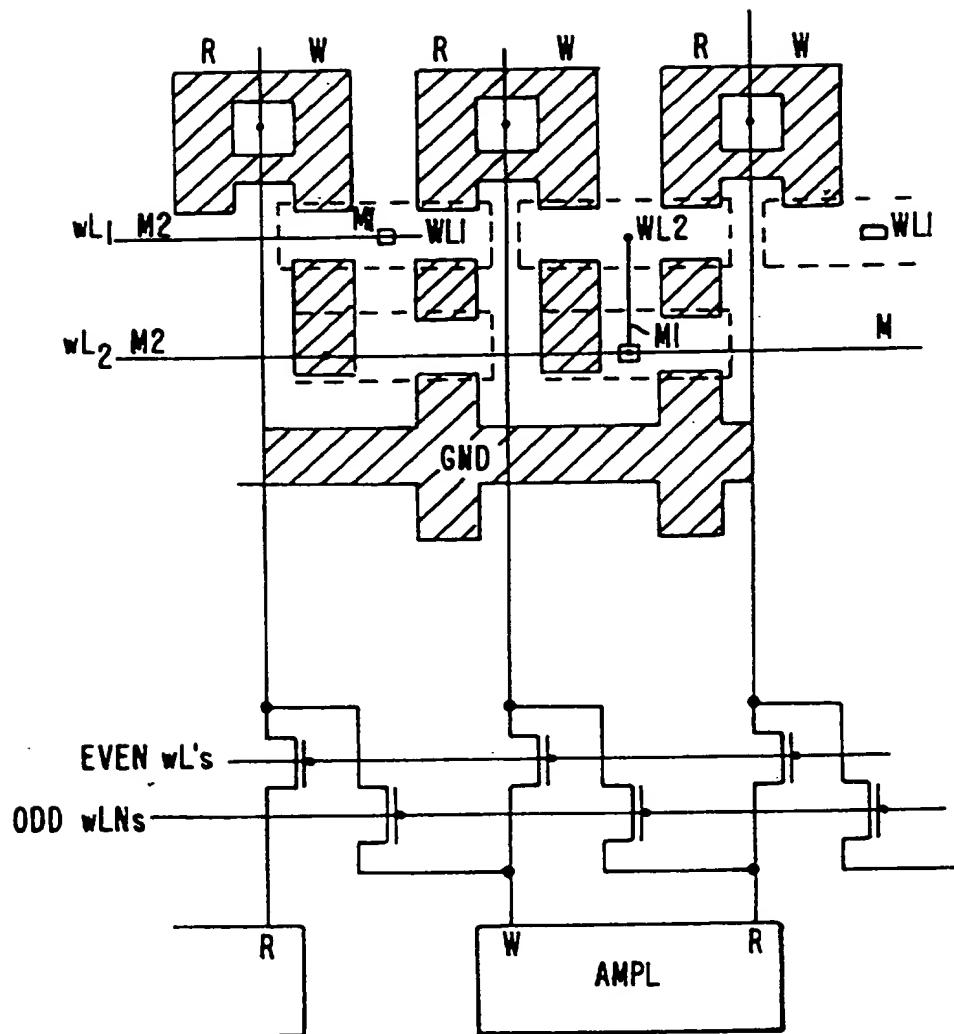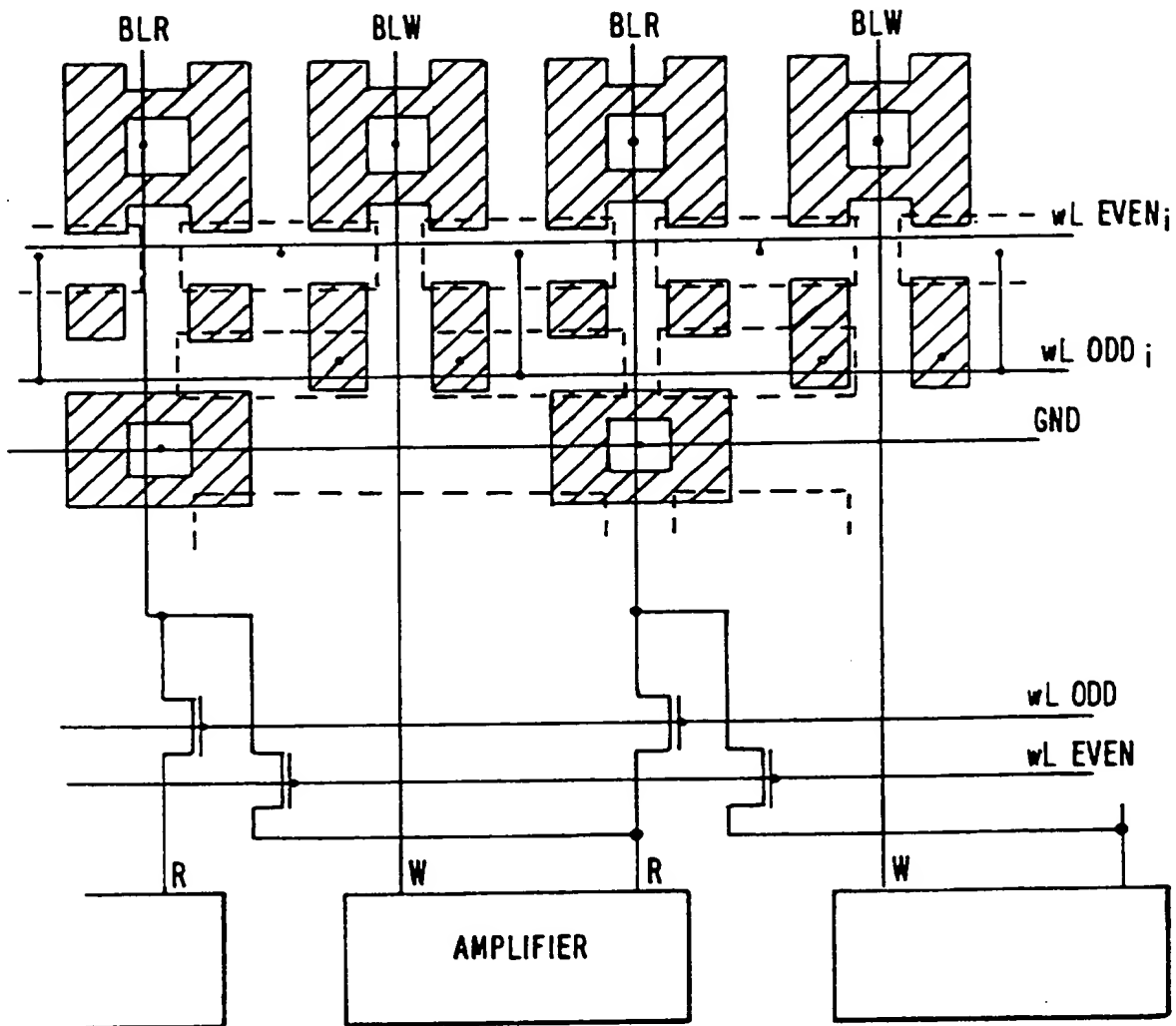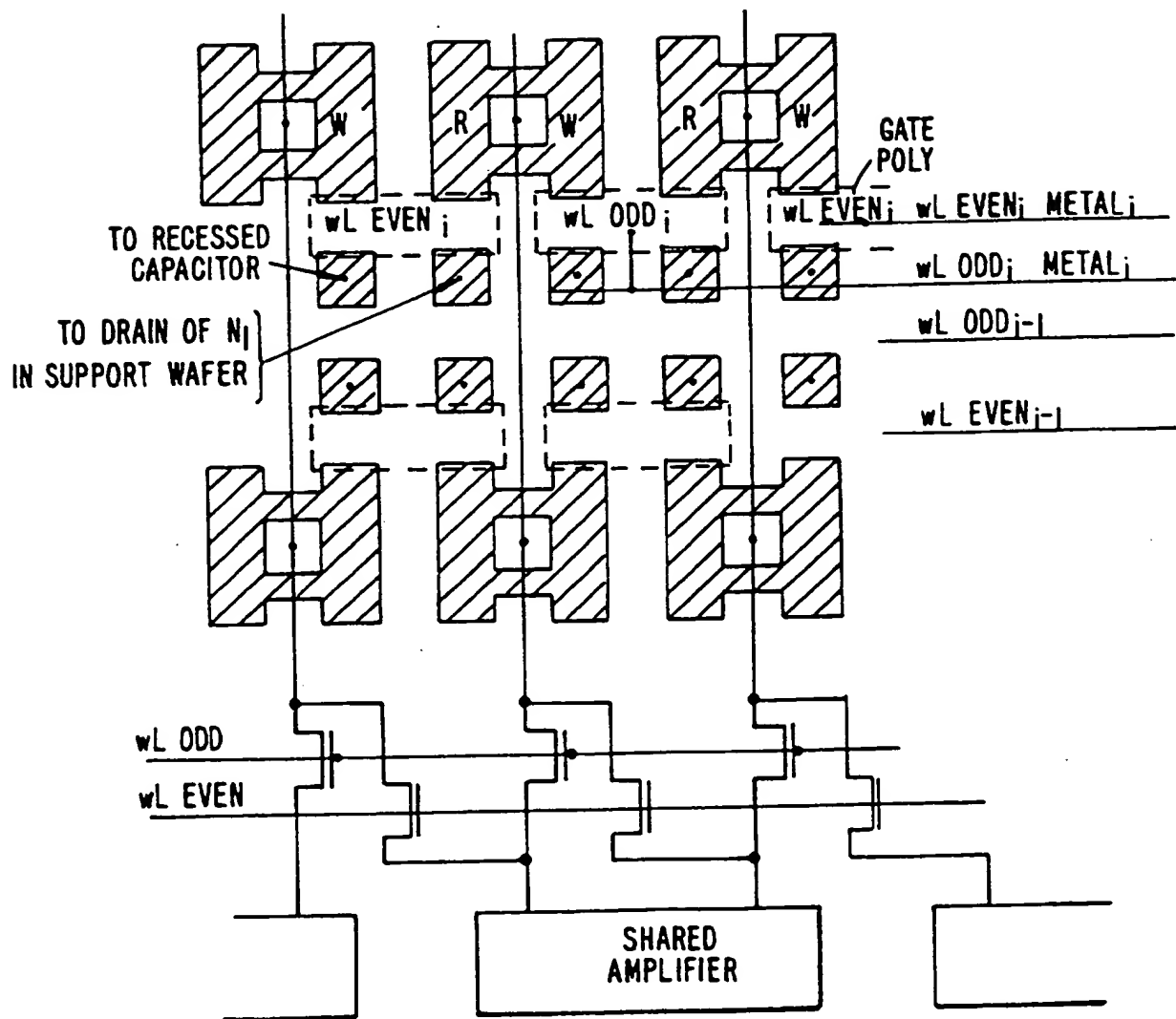
FIG. IIA.

FIG. IIB.

FIG. IIC.
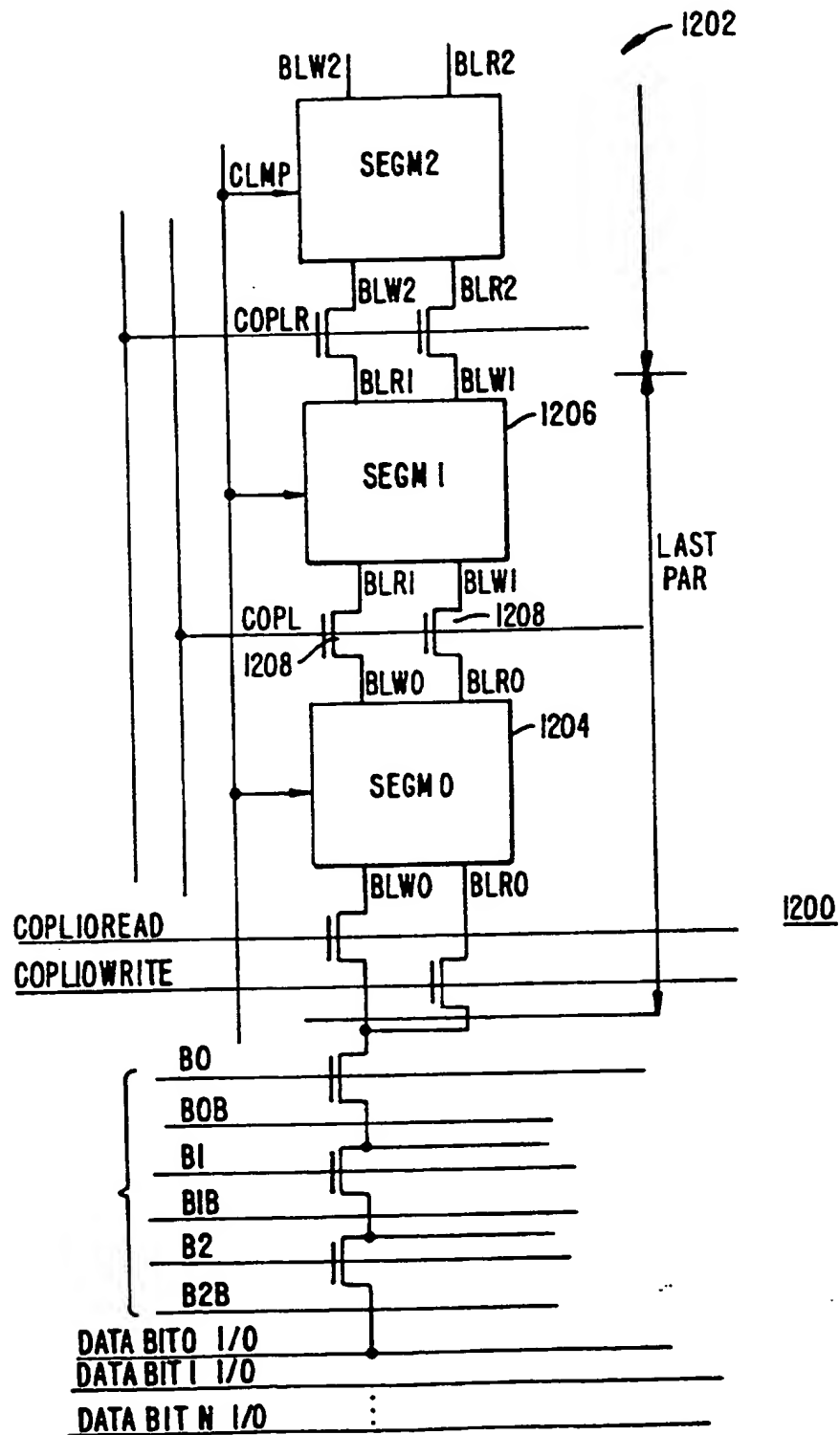
FIG. 12A.
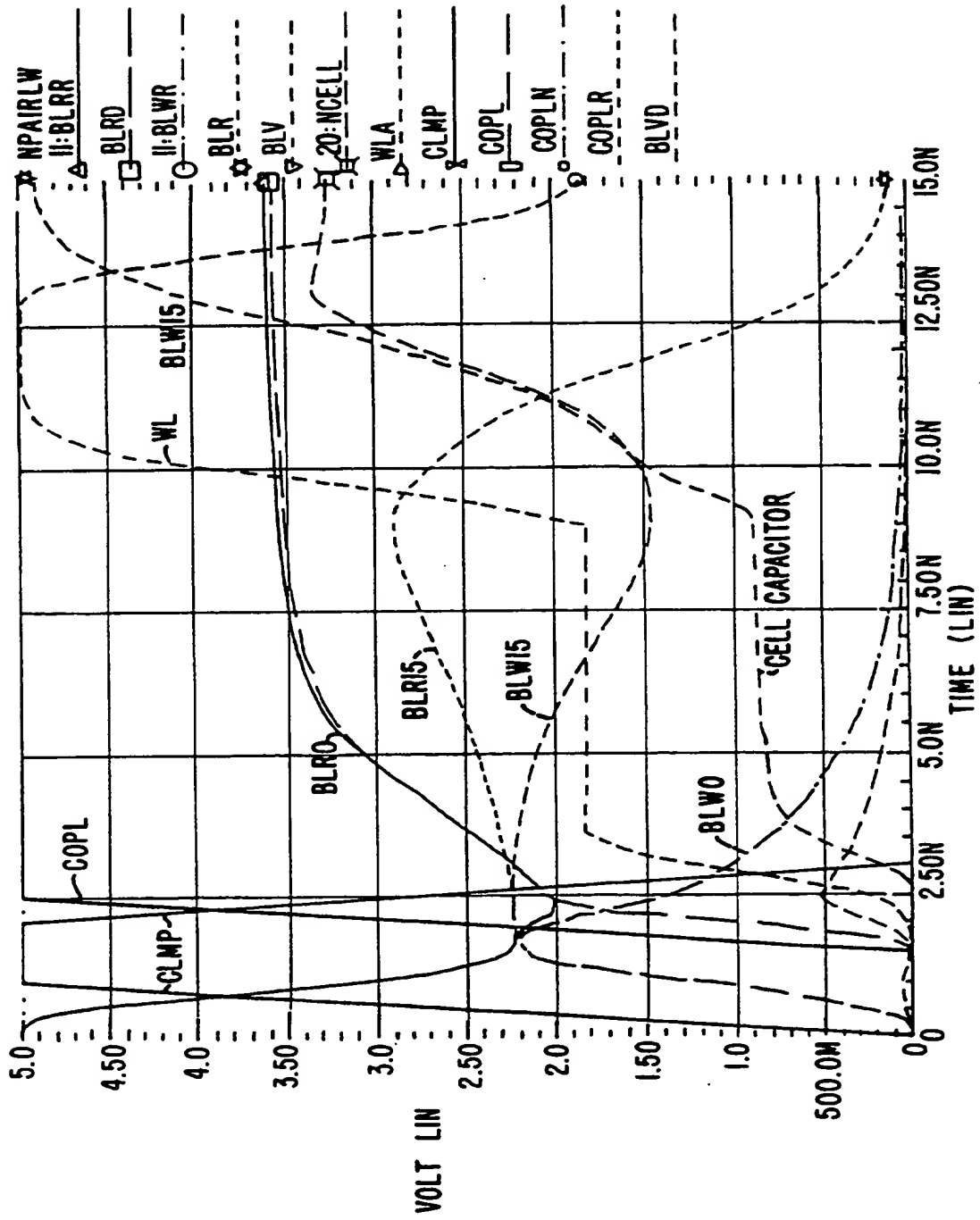
FIG. 12B.
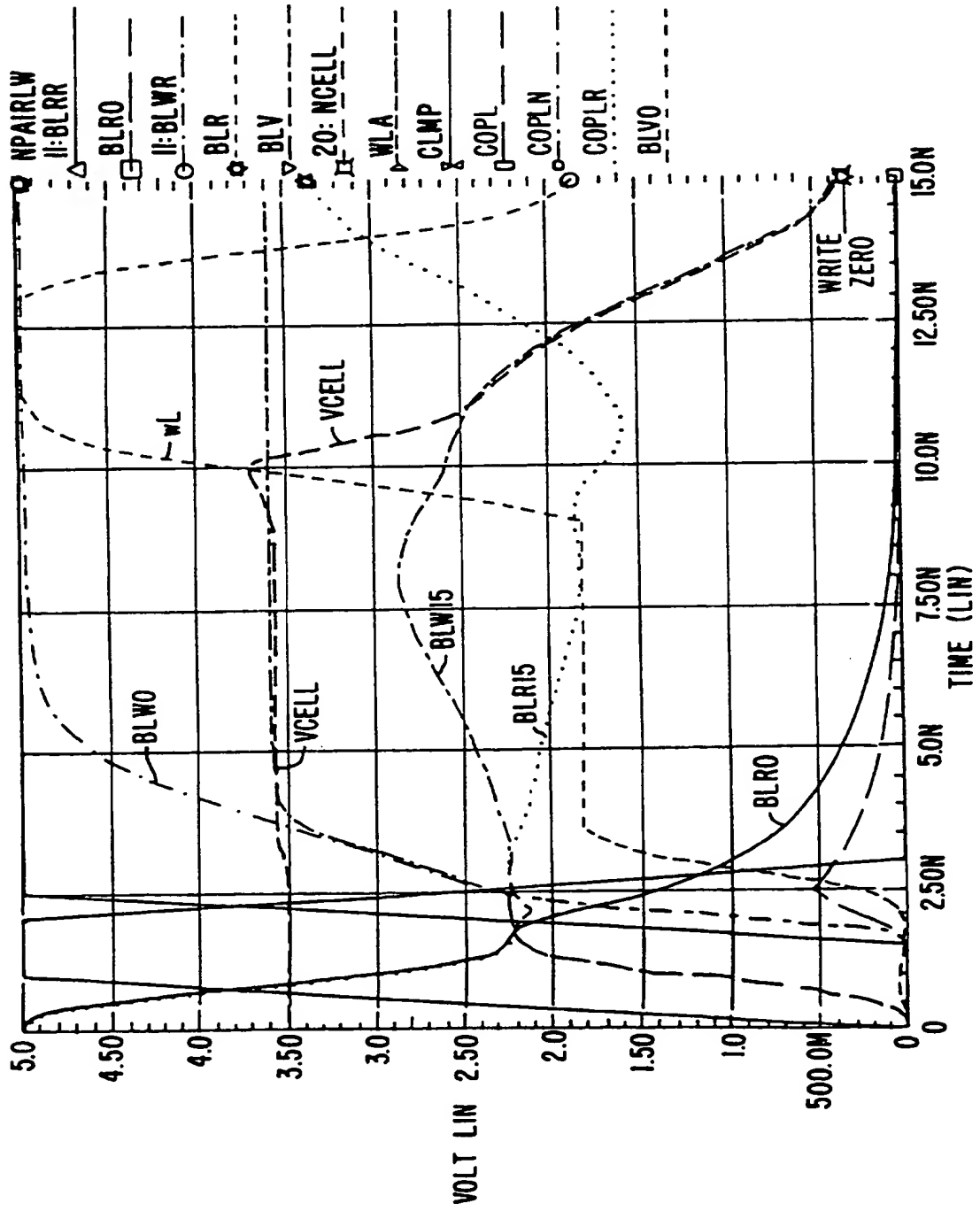
FIG. 12C.
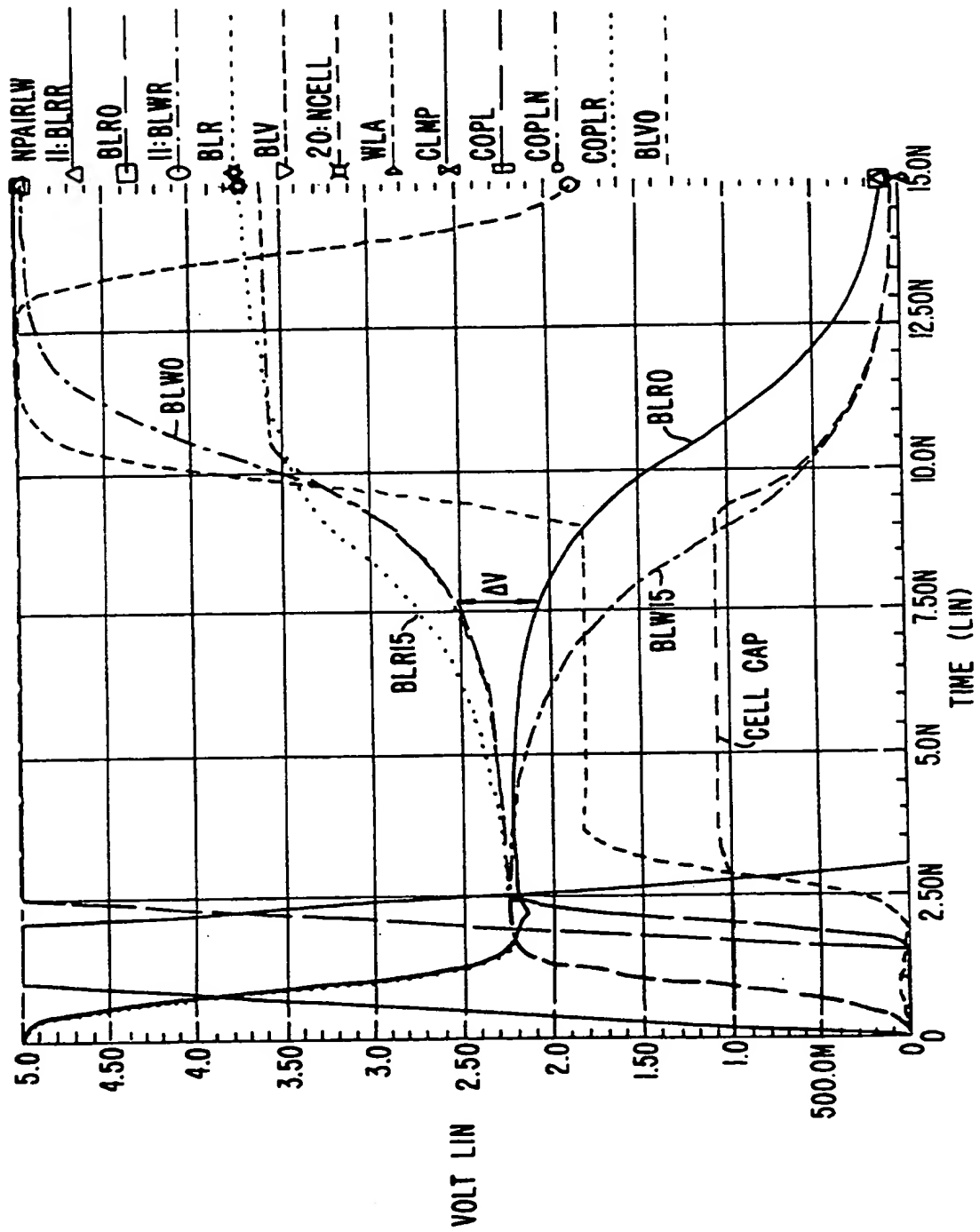
FIG. 12D.
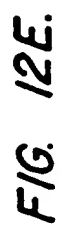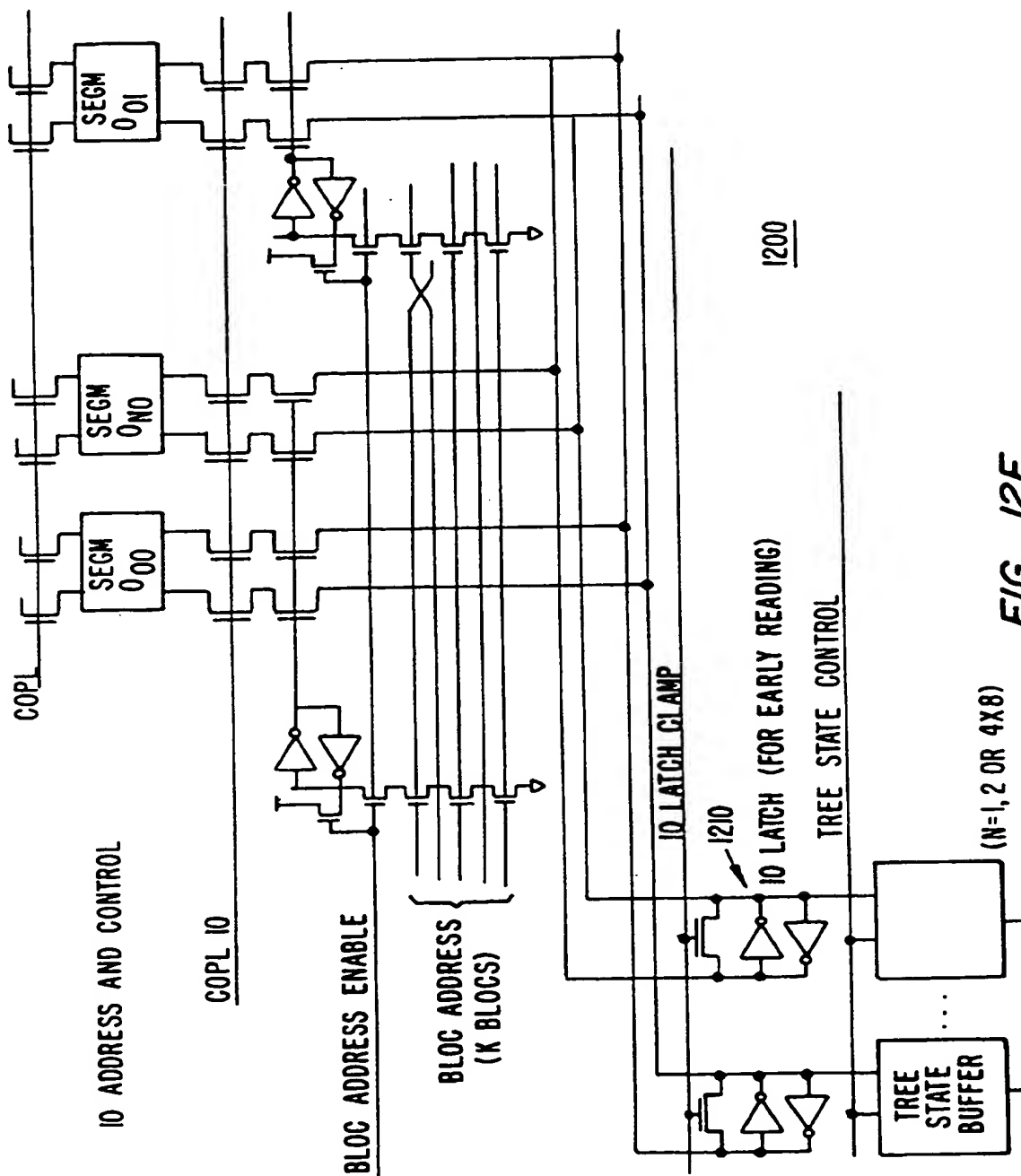
FIG. 12E.
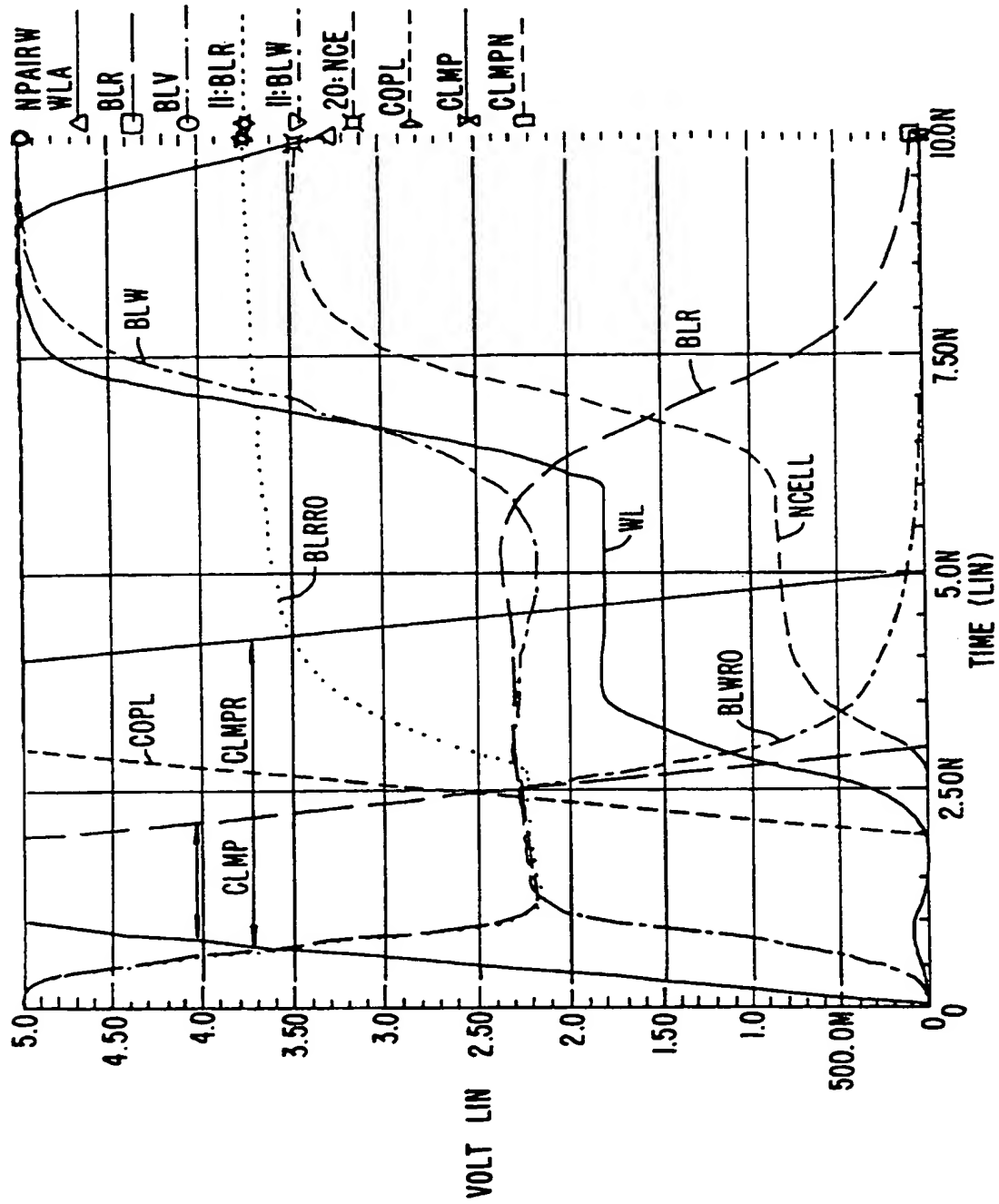
FIG. 12F.

FIG. 12G.

FIG. 12H.

FIG. 12I.

FIG. 12J.

FIG. 12K.

FIG. 12L.

FIG. 12M.

FIG. 13A.

FIG. 13B.

FIG. 13C.

FIG. 13D.

CLAMP

COPLR — COPL

COPLR —

COPLR AT
READ AND
WRITE

COPLR AT
AT REFRESH

BLR 1

BLW2

BLW 1

BLR2

VWL

STORED ZERO

POWER ON
SELECTED SEGMENT

VDDA

VSSA

POWER ON
NON SELECTED SEGMENT

VDDAR

VSSAR

FIG. 13E.

63/93

BLI   BL2

LOOKUP
TABLE
DECODER

WLBI

CONTROL NODE

GND

CONTROLLED
TRANSISTOR

CONTROL NODE

WLBO

DATA BUS

WLAI

CONTROL NODE

PASS
TRANSISTOR
DATA BUS

WLA O   CONTROL NODE

DUM I

VDD

DUM O

CLAMP

CLAMP LEVE

VDDA

1340

VSS A

COUPL

NEXT
SEGMENT

FIG. 13F.

FIG. 13G.

$(SEGM_i + SEGM_{i-1} + SEGM_{i-2} + \cdots + SEGM_0)$

CONTROL OF S

CLMP

$\overline{REFR}$

CLMP

COPLL

SEGMi

COPL

COPLE

SEGMENT DECODE

FROM SEGMENT i-1

$(SEGM_{i-1} + SEGM_{i-2} + \cdots SEGM_0)$

FIG. 14A.

ETC

SEGM4 SEL ——————————————— CONTROL OF
                                SEGM 4

SEGM3 SEL ——————————————— CONTROL OF
                                SEGM 3

SEGM2 SEL ——————————————— CONTROL OF
                                SEGM 2

SEGM1 SEL ——————————————— CONTROL OF
                                SEGM 1

$SEGM_0$ SELECTED

FIG. 14B.

FIG. 14C.

BLR   BLW

WLR

ONE
OF
N CELLS   WLW

CONTROL
NODE

CLAMP

VDD

AMPLIFIER

VSS

VDUMMY

COUPLING

TO
NEXT SEGMENT

FIG. 15.

BLR   BLW

CELL
MR   MW   WL
VWL

MS
C

VCLAMP   CLAMP
VDD

AMPLIFIER

VSS
VDD

VDUMMY

BLR
RESTORE
BLW   ZERO
C

C   BLW
RESTORE
ONE
BLR

DATA
IN/OUT

WRITE   READ
TIME   TIME

FIG. 16A.

FIG. 16B.

FIG. 16C.

FIG. 16D.

FIG. 17A.

FIG. 17B.

FIG. 17C.

FIG. 17D.

FIG. 17E.

FIG. 17F.

FIG. 17G.

FIG. 17H.

FIG. 17I.

FIG. 17J.

FIG. 17K.

FIG. 17L.

FIG. 17M.

FIG. 18A.
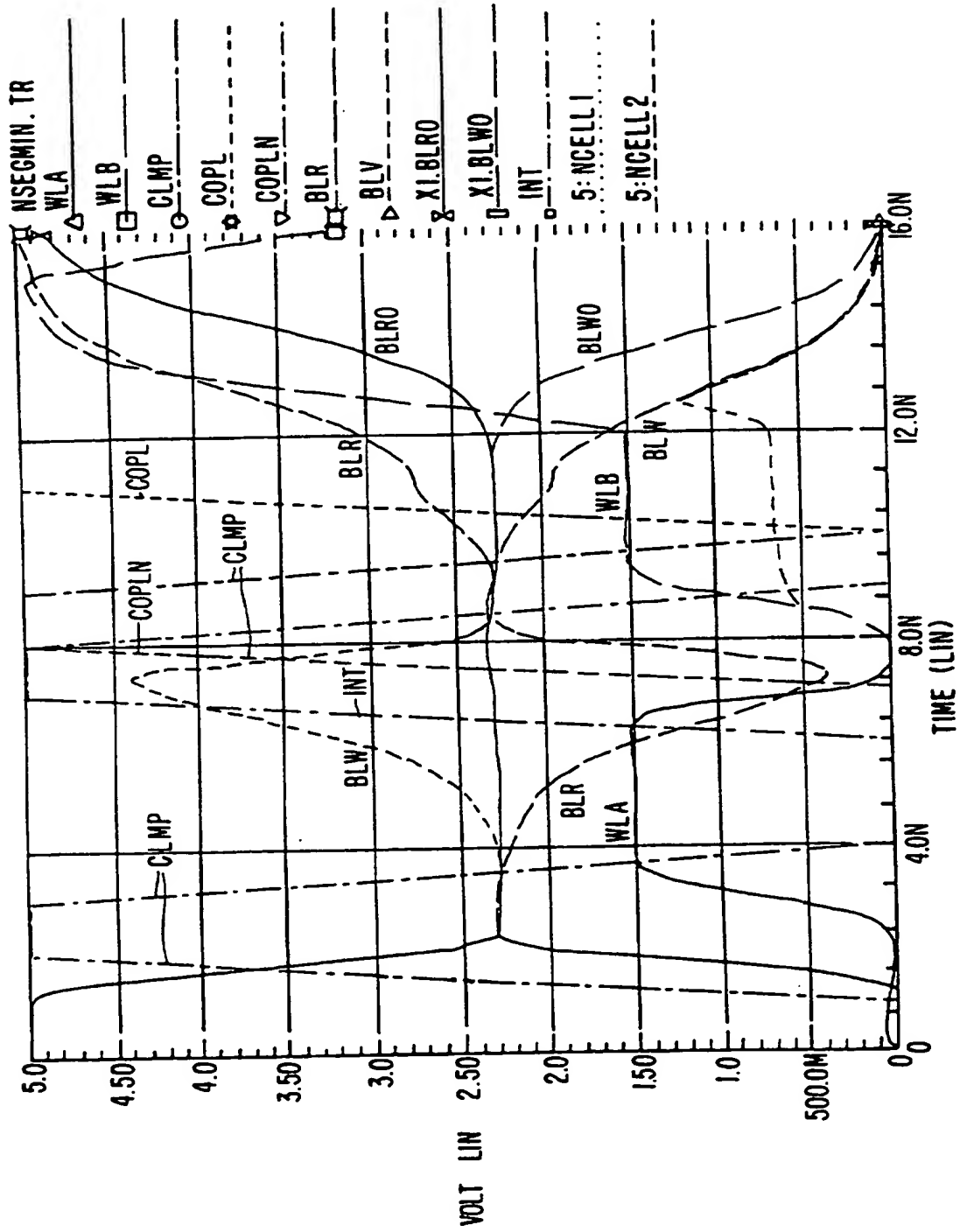
FIG. 18B.
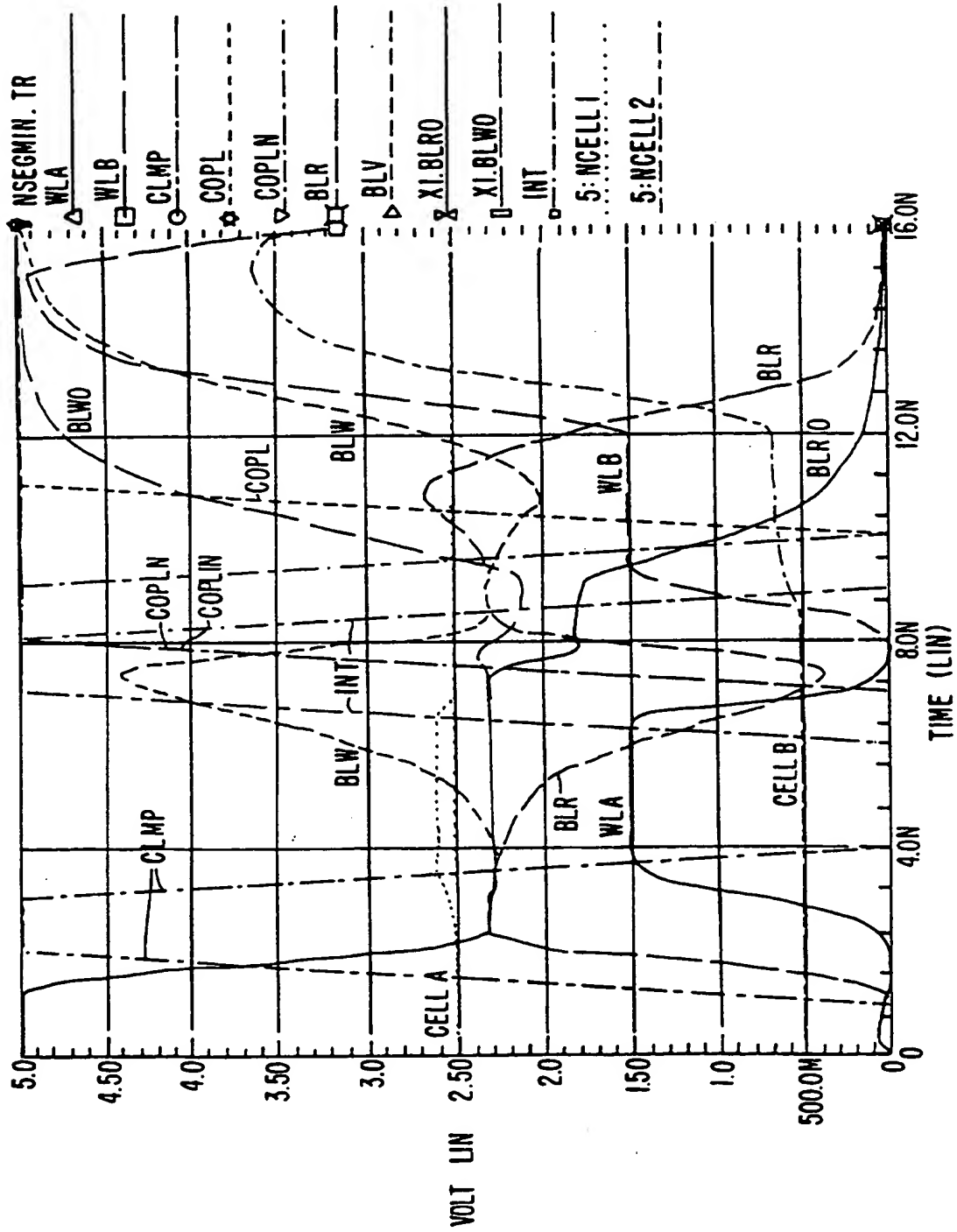
FIG. 18C.

FIG. 18D.
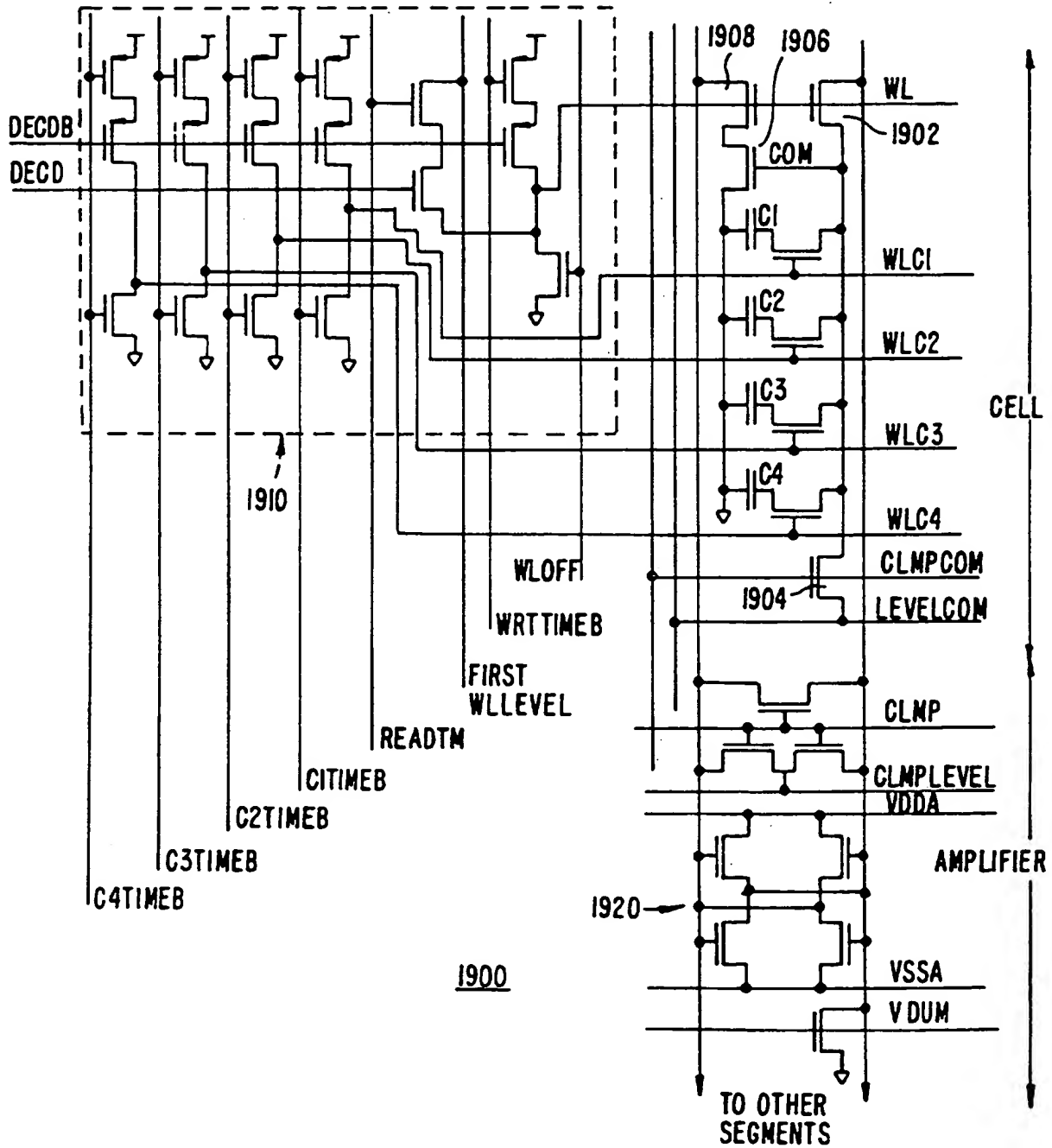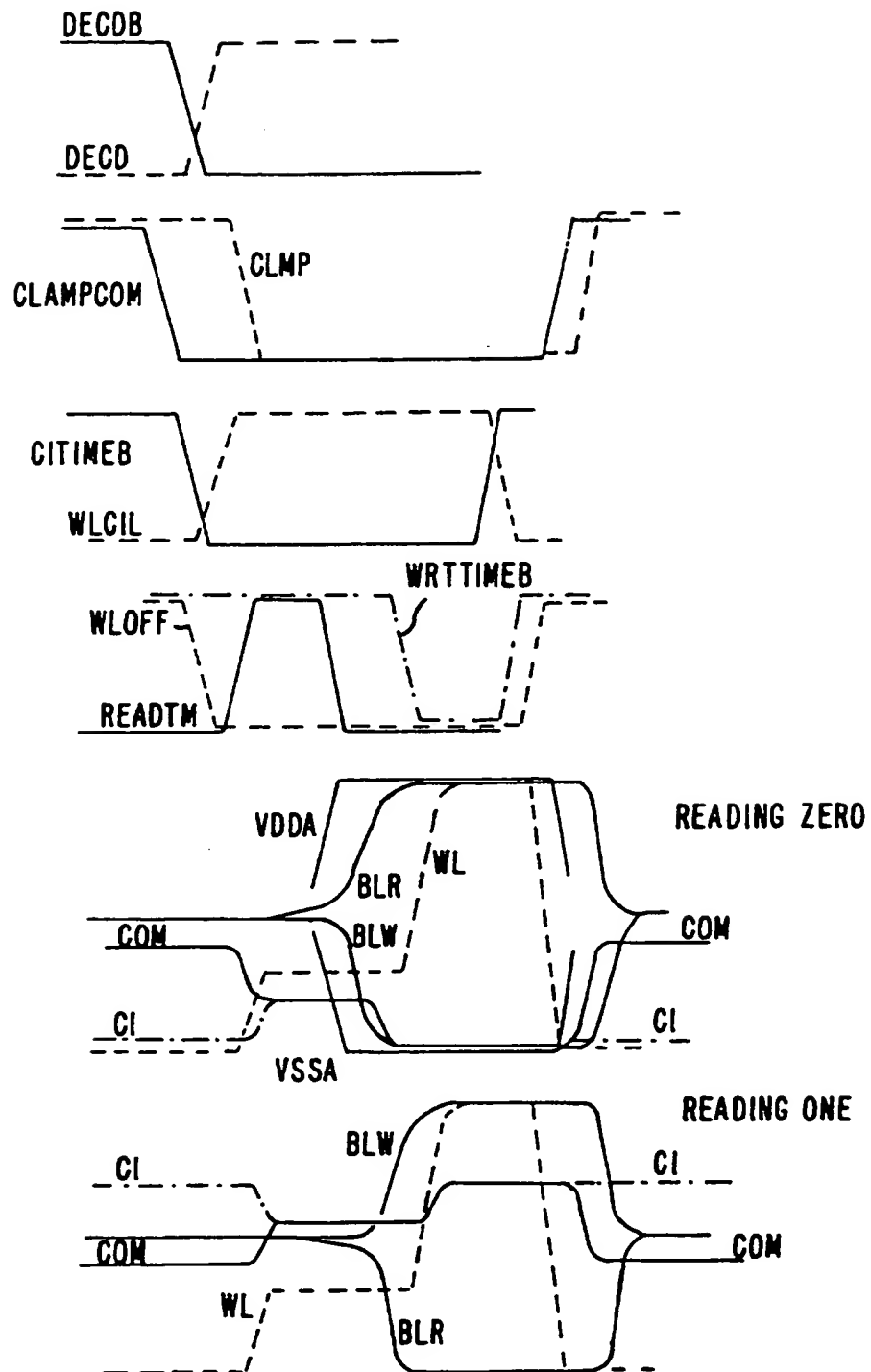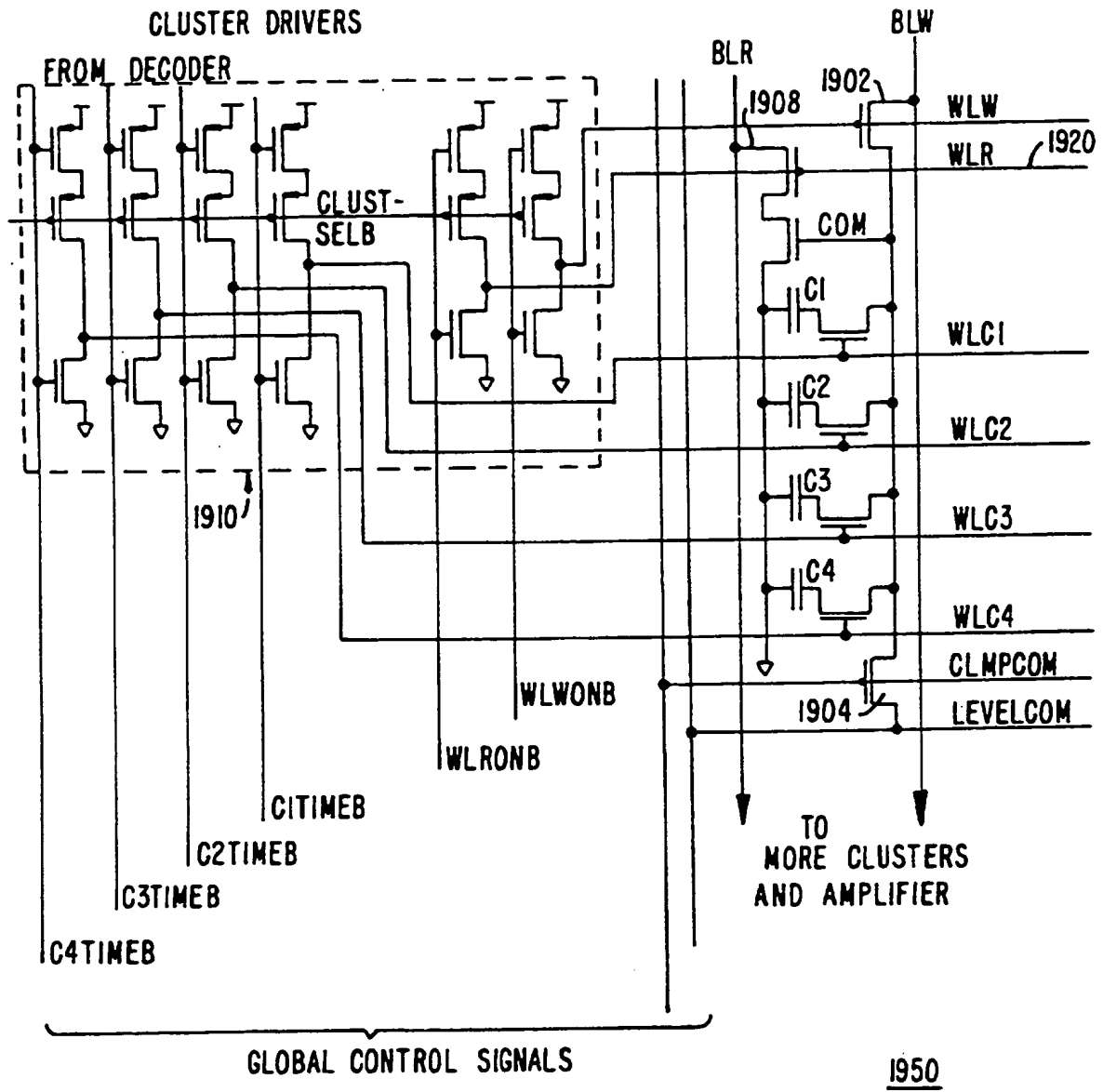
FIG. 19A.
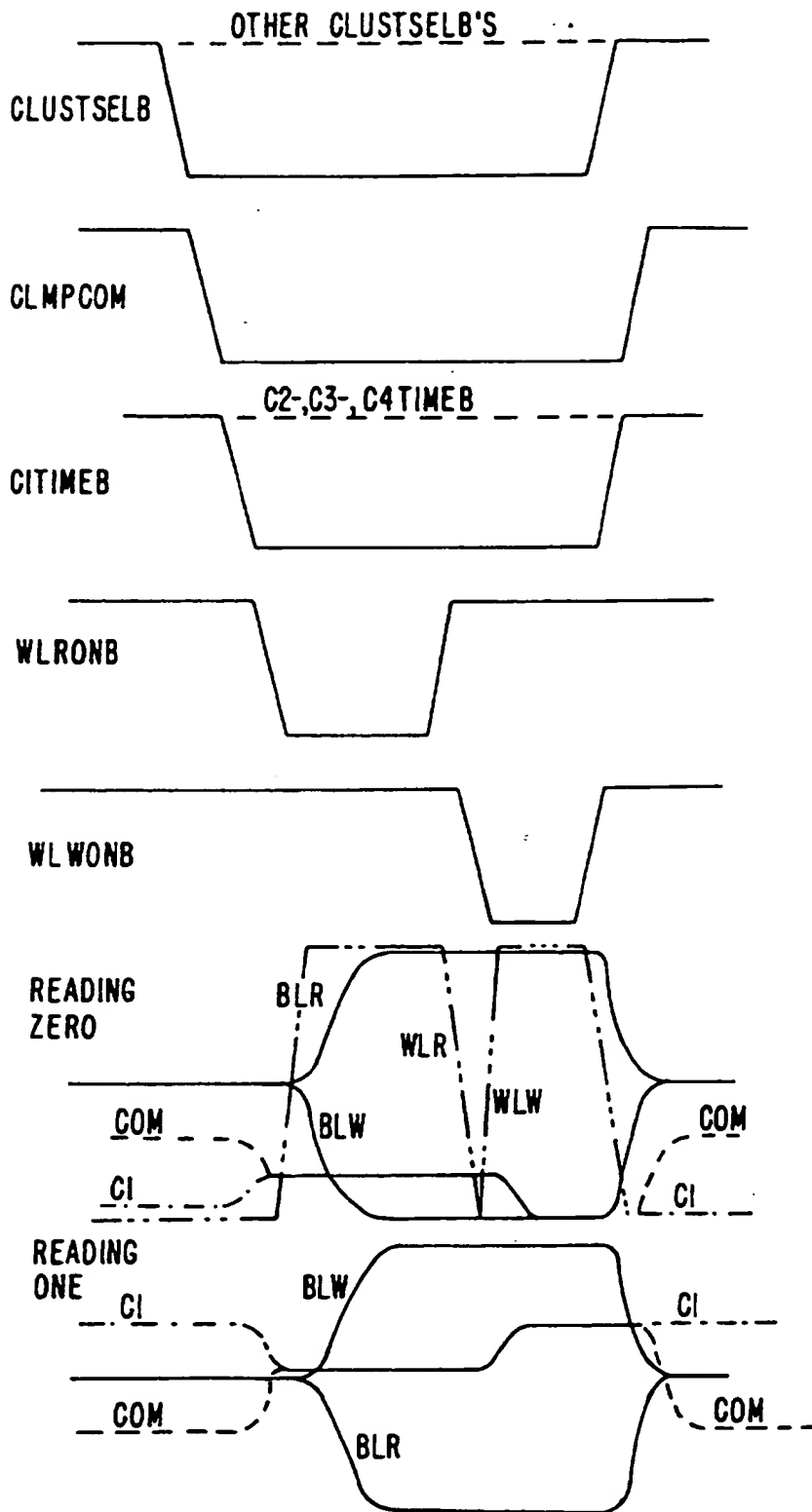
FIG. 19B.

FIG. 19C.

FIG. 19D.

DYNAMIC RANDOM ACCESS MEMORY

5

BACKGROUND OF THE INVENTION

The present invention relates to the operation and manufacture of integrated circuits. More specifically, in one embodiment the invention provides improved dynamic random

10 access memories (DRAMs), methods of operating dynamic random access memories and methods of making dynamic random access memories.

In the attempt to increase the number of bits on present DRAMs, methods are sought for shrinking device

15 dimensions while still maintaining high enough capacitance in the storage capacitors so that data can be reliably stored, refreshed and read. For example, new dynamic random access memories are disclosed in Application Serial No. 08/353,788, filed December 12, 1994, and U.S. patent number 5,396,452,

20 each of which is hereby incorporated by reference, wherein a memory cell contains separate read and write transistors.

In another example, substantial space could be saved merely by providing smaller capacitors in present DRAMs. However, smaller capacitors are generally refreshed more often

25 than larger capacitors, as it can be assumed that the leakage current is the same. The capacitance of a bit line is proportional to the number of bits on the bit line. The refresh power consumption is proportional to the frequency of refresh cycles multiplied by the number of bits per bit line,

30 where the refresh frequency is inversely proportional to the cell capacitor size. Therefore, the space advantages of a DRAM with smaller capacitors are mitigated by the higher power consumption of the required additional refresh cycles.

From the above examples it is seen that an improved

35 dynamic random access memory is needed, along with improved methods of operating such memories and improved methods of making such memories.

SUMMARY OF THE INVENTION

The present invention is directed, in one embodiment, to a memory structure having short bit line segments. Each bit line segment is coupled to a separate block of memory cells and a corresponding amplifier. The bit line segments are separated by pass transistors. The amplifiers are activated in all three modes of operation: read mode, write mode and refresh mode, while the pass transistors are enabled only in connection with data input and output. Very small cell capacitors can be used in this configuration, making it possible to use conventional gate capacitors, but power consumption is not appreciably increased. Furthermore, the speed of the memory for read and write operations is faster than present DRAMs with long continuous bit lines coupled to a single amplifier.

In another aspect of the invention a memory is provided wherein the contents of the memory can be read without any interference from ongoing writing, reading and refreshing. The memory is structured in two tiers, meaning that in addition to a first tier, a DRAM with addressing, reading, writing and refreshing, a second tier with separate addressing from the first tier is used to read the contents of the cells in the first tier.

Improved methods of forming a dynamic random access memory are further provided according to the present invention. For example, in one aspect of the invention, a method is provided for forming a memory cell for use with programmable logic devices that must be controlled with relatively large capacitors. In programmable logic devices, pass transistors transfer signals between locations in the device. The controlling capacitor must be significantly larger than the gate capacitance of the pass transistor, so that the voltage on the controlling capacitor is relatively constant during the signal transition. A multi-layer approach to forming the memory allows the controlling capacitor to be located underneath the pass transistors and the memory cell transistors.

A further understanding of the nature and advantages of the inventions herein may be realized by reference to the remaining portions of the specification and the attached drawings.

5

BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A-B show simplified block diagrams of a segmented bit line memory.

Figs. 2A-E show circuit and timing diagrams of latching and inverting amplifier circuits.

10

Figs. 3A-J show circuit diagrams of clamp circuits according to the present invention.

Figs. 4A-C show circuit and timing diagrams of a two inverter amplifier.

15

Figs. 5A-B show simplified block diagrams of control logic circuitry for the memory of the present invention.

Figs. 6A-C show circuit and timing diagrams of a single inverter amplifier.

Figs. 7A-E show circuit and timing diagrams of alternate embodiments of amplifier and control circuits according to the present invention.

20

Figs. 8A-H show circuit and timing diagrams of a two-tier memory.

Figs. 9A-F show process flow and layout diagrams for wafer preparation of the memory according to the present invention.

25

Figs. 10A-E show alternate process flow and layout diagrams for wafer preparation of the memory according to the present invention.

30

Figs. 11A-C show layout diagrams of a shared bit line memory.

Figs. 12A-M show circuit and timing diagrams for refresh read and write cycles in the memory of the present invention.

35

Figs. 13A-G show circuit and timing diagrams of control circuitry for the memory according to the present invention.

Figs. 14A-C show the disabling of segments above selected segments.

Fig. 15 shows DRAM cells with separate read and write word lines and a clamped amplifier.

Figs. 16A-D show memory cells according to the present invention with a full latch amplifier in each segment.

Figs. 17A-M show the result of simulations using a 0.5u NMOS process.

Figs. 18A-D show how an ongoing refresh cycle can be interrupted at any time during the cycle without destroying the stored data.

Figs. 19A-D show a DRAM having more than one storage capacitor associated with a single pair of read and write transistors.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Fig. 1A shows a simplified block diagram of a memory 100 including segmented bit lines 102, made up of a read bit line BLR and a write bit line BLW. Each bit line segment 104 is coupled to a separate block of memory cells 106 and a corresponding amplifier 108. Bit line segments 104 are separated by pass transistors 110 in line with bit lines 102. Amplifiers 108 are activated in all three modes of operation: read mode, write mode and refresh mode. Pass transistors 110 are enabled only in connection with data input and output. Address lines 112 select the particular memory cells 106 to be accessed. A word line decode block 114 is used to select a desired word line 116, while segment decode block 118 is used to selected a desired bit line segment 104.

Low order address bits select one of the word lines 116 in the segment selected by a segment decoder. The higher order address bits select a bit line segment 104. In a refresh mode, the low order bits select one word in each segment, and the segment decoder is bypassed to enable all segments. It should be noted that in an alternative embodiment, rather than separating the address lines into low and high order bits, a single set of multiplexed address lines may be implemented. The bit lines in the selected segments

that are normally clamped to a clamp level of Vdd/2 may be released to float during a period when the clamp is off. The selected word line 116 is then brought to the first ramp level, and a dummy current is turned on for the selected bit
5 line segment 104. The voltage on the read bit line in the segment changes quickly due to the low bit line capacitance. The voltage displacement due to the cell current can actually be of the same order of magnitude as the final displacement caused by the amplifier, when its power is turned on later in
10 the refresh cycle. The amplifier 108 will therefore use most of its power dissipation to bring the bit lines to the power rails. Both latch type and the inverter type amplifiers, each of which will be described further herein, can be used. As described earlier, the ramp is brought to its high level for a
15 moment, increasing voltage on word line 116, when the high bit line has reached Vdd, completing the refresh operation before power again is turned off. In the very end of the cycle the clamp signal is applied again.

In order to read data from memory 100, all bit line
20 segments 104 are initially disconnected, and the clamp input is brought low on all segments. This permits bit lines 102 in all segments to float. Only one word line 116 in only one bit line segment 104 is selected in the read mode. When the read cycle has reached a point where amplifier 108 has been turned
25 on and amplified the signal slightly, the bit line segments 104 on one bit line 102 are connected. The interconnected segment lines initially appear as delay lines consisting of RC elements. The capacitance C corresponds to the capacitance of the individual bit line segments 104, while the resistance R
30 corresponds to the impedance of the coupling transistor. This is however not a regular delay line, as each segment 104 also has its own amplifier 108. This allows a signal to proceed from segment to segment very quickly while also being amplified. In a typical memory configuration, a column
35 decoder would select the columns to be the subject of the reading or writing of data. The non selected columns will still be addressed by the word line 116 selected by the word line decoder 114. When the amplitude on the output segment

has reached the rails, the segment is connected by the column decoder to the output register of the memory.

In a write operation, data is connected via the column decoder to the bottom segments of the selected columns. The write circuit forces the bottom segment separation to be much larger than the separation of the segments in the addressed and read segment at the time of coupling the segments together. If the read data is different than the data to be written, two signals of different polarity will proceed in opposite directions down the "delay line". The signal with the largest amplitude will determine the final polarity of the addressed segment.

In the refresh mode, the segments remain disconnected during the full read-restore cycle. As mentioned above, one word line in each segment is selected by the low order address bits in the word line decoders and the segment decoder is bypassed so that all segments are activated the same way. The dummy current circuit is also activated in all segments. The control circuit in Fig. 1B clearly shows how the refresh mode differs from the read and write modes in the word line decoding. It should be noted that if the amplifier 108 in each segment 104 consists only of an inverter, the time required for refresh may be longer than the time required for reading and writing. The read bit line in each segment is driven only by the difference of cell current and dummy current. This current is increased significantly at the end of the cycle, but the signal on the read bit line is not enhanced by amplifier 108. In reading and writing, on the other hand, segments 104 are interconnected, so that the read bit line on the first segment is connected to the write bit line of the second segment and so on. This means that the output of one segment amplifier drives the input to the next segment amplifier in the read and write modes. The delay through the "delay line" must of course be taken into account when comparing the timing requirements. It is preferred that all modes have the same timing, as this simplifies the global control circuits.

Therefore, very small cell capacitors can be used in this configuration, making it possible to use conventional gate capacitors. The speed of the memory for reading and writing is faster than the same type of memory with long

5 continuous bit lines.

The addressing and control of memory 100 is shown in Fig. 1B. Memory sequence generator 140 receives memory control signals such as Memory Request and Read/Write Enable and provides amplifier control signals 142. Amplifier control

10 signals 142, combined with high order segment address bits 112, are input to segment decoder control block 118 (Fig. 1A) to control a selected one of memory segments 104. Refresh counter 144 initiates a refresh cycle when necessary. The output of refresh counter 144, combined with low order word

15 address bits, are input to word line decoder block 114 (Fig. 1A) to select a particular word line in memory 100. Memory sequence generator 140 also generates additional control signals, such as Copl I/O, Block Address Enable and I/O Latch Clamp, that are input to a column decode block (not shown)

20 that accesses a desired column of memory 100 for reading, writing or refreshing.

The type of amplifier 108 used depends on the application for memory 100. For the highest possible speed, where disturbance on the cell capacitor voltage is permitted

25 to be high, a basic latching amplifier 200 coupled to memory cells 202, as shown in Fig. 2A, may be used. Memory cell 202 includes a storage capacitor 204, a write transistor 206 and a read transistor 208. Amplifier 200 includes clamping transistors 210 and cross-coupled inverters 212. The

30 corresponding timing diagram is shown in Fig. 2B. It should be noted that for highest speed with a given transistor size in memory cells 202, a higher initial cell current can be obtained if the threshold voltage on write transistor 206 is increased. There is of course a great advantage to using a

35 standard CMOS process for the memory, especially if mixed with logic. The segmented bit line approach and small cell capacitors 204 are preferred to maintain the highest possible

memory speed.   In that case, the segments would be coupled together just after amplifier 108 has been powered up.

In operation of the memory shown in Fig. 2A, at the end of a memory cycle, the power to the amplifier is turned off and VCLAMP is turned on, clamping the bit lines to the clamp voltage level (VCLAMPL), which is typically about half of the Vdd level. At the beginning of a new memory cycle, VCLAMP is brought low, allowing the bit lines BLR and BLW to float. VWL is then brought to an intermediate level that is at least high enough to draw current from the read bit line BLR if the voltage on cell capacitor 204 is high (a stored ONE). If, on the other hand, the voltage on cell capacitor 204 is low (a stored ZERO), an excessively high signal level on VWL will charge the cell capacitor through write transistor 206 during the read portion of the cycle. A slight increase of the cell voltage while reading a stored ZERO from cell capacitor 204 is acceptable since the resulting read current is minimal compared to the cell current while reading a stored ONE. At the same time the first word line voltage is applied, a negative going voltage is also applied to dummy transistor 214, causing it to feed a current to BLR which is designed to be about equal to the average current for reading a stored ONE or for reading a stored ZERO from cell capacitor 204. As shown in Fig. 2B, BLR is pulled in the positive direction when a stored ZERO is read because the dummy current dominates. While reading a stored ONE, however, the cell current dominates and pulls BLR in the negative direction. As soon as the voltage difference between the bit lines is sufficient for reliable operation, the supply voltages VDDA and VSSA for latching amplifier 212 are applied. Bit lines BLR and BLW will then quickly move to their respective rail voltages. Once each bit line is close to its final voltage, the voltage on the word line VWL is increased to at least a level of VDD. A previously stored ONE will now be refreshed to approximately one threshold voltage less than VWL, while a previously stored ZERO will be restored to 0 V.

If speed is not the primary goal and the disturbance level on the cell capacitors 204 is allowed to be high, the

basic inverting amplifier 220 coupled to cells 202, as shown
in Fig. 2C, may be used. The operation of the memory shown in
Fig. 2C is similar to that of the memory shown in Fig. 2A.
Amplifier 220 includes clamping transistor 210 and a simple

5   inverter 222. The clamping to a common reference level, which
is chosen to be close to the switching point of the amplifier,
is not very accurate, as the switching point of the amplifiers
varies with changes in the device parameters. The timing must
therefore be adequate to allow the cell current to displace

10   the read bit line BLR at least as much as the error before,
since the bit line voltages represent the actually stored
data. The power to the inverter amplifier 220 should not be
turned on before the worst case clamping error has been
compensated. This will prevent the write bit line BLW from

15   temporarily going to a ground level with the word line on,
which would cause a stored one to be discharged to a zero. A
stored zero on the other hand would not be affected if the
write bit line inadvertently went to the high level, while the
word line voltage is still at its lower level. The preferred

20   method is to wait a sufficient time until a potential clamping
error has been compensated for. The amplifier will then
consistently move the write bit line in the correct direction.
As discussed earlier, power consumption can be minimized if
power is turned on even later in the read mode, so that read

25   bit line BLR has moved closer to a rail voltage. If the
inverting amplifier is used in connection with segmented bit
lines, the coupling can be made only when the clamping error
has been compensated and the signal has been amplified
somewhat. Fig. 2D shows the timing for amplifier 220 in

30   Fig. 2C, and Fig. 2E indicates the amplifier's performance as
part of a segmented bit line configuration. Note in the
latter case how the partially amplified signal on both bit
lines BLR and BLW is temporarily reduced when the adjacent
segment is connected. Again, the read bit line BLR on the

35   read segment is connected to the write bit line BLW on the
next segment. The segments together act as a plurality of
latched amplifiers. It is important that all amplifiers are
powered up before the segments are coupled, as this guarantees

a gradual amplification of the initially reduced signal before
it progresses down the full bit line. Without amplification,
the signal on the read bit line BLR of the read segment would
be reduced to a small fraction of its original value. For a
5   memory structure where a bit line is made up of eight
segments, then, the signal would be reduced by a factor of
eight. This signal is too weak to overcome the clamping error
on some of the segments.

10   <u>Compensation of Clamping Error</u>
            In applications where the cell current is low, yet
fast memory cycle times are still needed, it is desirable to
minimize the clamping error. Figures 3A through 3G illustrate
the different clamping devices and methods. Fig. 3A shows a
15   preferred structure with clamping transistors 302 and 304,
which was also shown in Figs. 2A and 2C. Both bit lines will
reach the reference level VCLAMPL, while VCLAMP is high, but
when VCLAMP goes low, the gates of the transistors 302, 304
will couple a small negative charge to each of the bit lines.
20   The voltage drop on the bit lines can easily be as much as 100
mV, if a short clamp time is desired. To achieve a short
clamp time, the clamp transistors 302, 304 must be relatively
large. Segmented bit lines, with their lower capacitance,
will have smaller clamp transistors than full-length,
25   unsegmented bit lines, but the ratio of gate capacitance to
bit line capacitance is the same and therefore the result is a
comparable disturbance. The latched amplifier 200 (Fig. 2A)
is relatively immune to the clamping offsets that appears
equally on both bit lines. Latched amplifier 200 suffers from
30   a non-correctable offset error due to device parameter
variations. The two cross-coupled inverters 212 may have
different switching points. One of the bit lines will
therefore be moving at a different rate after clamping than
the other bit line. The cell current must be larger than the
35   built-in offset current for a correct reading of the selected
cell. If used in a segmented bit line configuration, there is
a risk that one or more of the non-selected segments may start
to switch on their own before the segments are connected,

which could result in a read error. The clamp circuit in
Fig. 3B addresses this possibility by adding transistor 306
that directly shorts the bit lines during clamping.

The inverting amplifier 220 (Fig. 2C) is sensitive

5   both to the difference between its switching point and the
reference VCLAMPL and to the effect of turning off VCLAMP.
Instead of waiting for the read bit line to overcome the worst
case clamp error, another approach may be taken, as shown in
Fig. 3C. VCLAMP1 again connects both bit lines to a common

10   reference voltage VCLAMPL while power is off, but VCLAMP1 is
turned off as soon as power is turned on. VCLAMP2, on the
other hand, turns on at the same time as VCLAMP1, but stays on
longer. When power is turned on, the bit lines are at the
potential of VCLAMPL, which may be different than the

15   switching point of the inverter. VCLAMP2 shorts the input and
the output of the inverter, so that the bit lines are moved to
the switching point, thus putting the inverter in full balance
before the cell current is applied to the read bit line. Note
that turning off VCLAMP2 causes a negative displacement of the

20   bit lines. The displacement on the read bit line (the
inverter input) is most critical and must be compensated by
coupling a positive charge to the read bit line. The clamp
shown in Fig. 3D accomplishes that objective. Here, the
connection to the reference voltage is done with a P-channel

25   transistor 308. VCLAMPB goes positive at essentially the same
time that VCLAMP goes low. The P-channel transistor 308 has a
relatively high on resistance compared to the N-channel
clamping transistor 306 and the impedance in the inverter
associated with searching the switching point centering. The

30   purpose of the P-channel transistor 308 is to maintain the bit
lines close to the switching point at for a longer stand-by
period. Since the switching point is set to half the Vdd
level (Vdd/2), and the bit lines before clamping are set to
the rail voltages, the potential of the bit lines after

35   clamping will therefore be close to Vdd/2. The width and
length of the P-channel transistor 308 can be chosen such that
the requirement for both high impedance and the required
coupling of charge to the read bit line BLR can be met.

Fig. 3E illustrates the addition of compensating capacitors 310 and 312 between VCLAMPB and bit lines BLR and BLW that are chosen to fully compensate at turnoff of the clamp transistor 308. In Fig. 3F, VCLAMP1 holds bit lines BLR and BLW at the reference level in standby mode, while VCLAMP2 remains high after power has been turned on to inverter 220 (Fig. 2C). The charge transfer to the bit lines, when CLAMP2 is turned off, is compensated by VCLAMPB going positive. The compensation is necessary on the inverter input, but is less critical on its output, so the compensation can be limited to only one capacitor on the read bit line.

As will be discussed herein, the advantageous features of inverting amplifier 220 and segmented bit lines in Fig. 1A can also be applied to conventional DRAM cells. In that case, the DRAM cells are connected to both bit lines. This type of clamp circuit is shown in Fig. 3G. The associated amplifier includes two separately powered inverters and two separately driven compensation capacitors 314 and 316. A similar amplifier arrangement can also be used for the DRAM cells of this invention, except that a common drive signal for the compensating capacitors shown in Fig. 3F can be used. The compensating capacitors used in a conventional DRAM may serve two purposes. First, they supply the charge to compensate for the negative charge transferred when VCLAMP2 goes negative, but the size of the capacitor is a little larger than required for that compensation. The added charge pre-biases the bit line to which the read cell is connected, so that reading a cell will give the same absolute voltage difference between the read bit line and the switching point for both a charged and a discharged cell capacitor. In reading and writing data in a segmented bit line configuration, the segments that are not addressed must have their read bit lines BLR very close to the switching point not to cause errors. The compensating charge transfer for these segments, must therefore be less than for the selected segment. This can be accomplished by using a lower amplitude on the compensating voltage or, of course, a separate compensating capacitor can be used to add the dummy charge on the selected segment.

The amplitude of the clamp signal for all the cases discussed above is the full Vdd voltage, resulting in a large capacitive coupling when the clamp signal is turned off. In the initial time period after clamping, the bit lines will

5 still be close to the clamping level (Vdd/2). This means that the clamp signal need not initially go more negative than to ((Vdd/2)+Vth). As the cell current proceeds to separate the bit lines, the clamp signal must track the negative-going bit line. The arrangement in Fig. 3H shows such an configuration.

10 Here, when VCLAMP is turned on, transistor 318 lifts the gate of transistor 320 to (Vdd-Vth), which clamps the bit lines together. The transistors 322 and 324 act as diodes and are designed as relatively weak transistors in order to limit how much they can pull the clamped bit lines positive while VCLAMP

15 is high. Transistors 322 and 324 are actually designed so that the positive excursion during the time of clamping is exactly as much as the negative capacitive coupling when VCLAMP goes low. When transistor 318 is turned off, transistors 322, 324 proceed to pull the gate of clamp

20 transistor 320 toward the lowest bit line voltage plus the threshold voltage of transistor 322 and transistor 324. With a normal threshold, this tends to hold a slight clamping for a short period, reducing the gain of the inverter for small cell currents. By terminating the clamp signal early or by using

25 lower threshold voltage on 322 and 324, this problem can be eliminated. When the bit lines eventually move, the low-going bit line will continue to keep the gate of the clamp transistor 320 just below its threshold voltage.

In several of the amplifiers, the assumption has

30 been made that the amplifier power is permanently connected. This means that the bit lines are close to Vdd or Vss between cycles. A clamping operation will then have to swing the bit lines as much as Vdd/2 to the switching point of the amplifier. The time required for the clamping operation is in

35 this case slightly longer than when the bit lines start out close to the clamp level. The clamping circuit as shown in Fig. 3I is used for this purpose and consists of one N-channel clamp transistor 326 and two capacitors, here shown as

transistors 328 and 330 with both source and drain connected to the associated bit line, for compensation of the turnoff disturbance.

Fig. 3J shows a single N-channel clamping transistor 332, used in connection with clamping under power. Amplifier power is selectively turned on and off during a memory cycle. In cycles repeating frequently, the bit lines start out very close to the switching point. At the end of each cycle the two bit lines are clamped together after having been close to the opposite rail voltages. The refresh frequency must satisfy the requirements for the small cell capacitors, so larger bit line capacitances will discharge very little between cycles.

The amplifier's VddA and VssA nodes are both connected to the reference level Vdd/2, which is the nominal switching point level. Any leakage through the transistors would go to this level. Only junction leakage could cause the bit lines to drift away from the reference level. The inverter transistors would however limit the drift to one threshold away from the reference level in either direction. Starting from this worst case situation, the adjustment toward the switching point after applying power to the amplifier follows the table below. The inverter has a W/L ratio for the N and P transistors of 1.2/.6 and 2.4/.6, respectively, while the bit line capacitances are 100 fF (100E-15).

Time after power on	VBLR - Vref
ns	mv
0	-600
2	-376
3	-163
4	- 68
5	- 19
6	- 5
7	- 1
8	0

Turning off a clamp transistor of with W/L of 2.4/.6 introduces a disturbance of -80mV. The disturbance must be considered in relation to the cell current. The rate of voltage change on the read bit line is 10mv/(μA*ns) for the bit

line capacitance of 100 fF. A differential cell current (Icell
- Idummy) of 5µA would change the bit line voltage at a rate of
50mv/ns. Turning off the clamp after 5ns, with the cell
current on, would result in full compensation 7ns after the

5   start of the cycle, overcoming 80mv plus 19mv. At lower cell
currents the clamping will be turned off about 1ns later, and
it would take longer to compensate for the offset caused by the
clamp transistor. At a differential cell current of 1µA, the
total time to compensation would be about 14ns. The polarity

10   of the disturbance on the read bit line is such that the write
bit line for the lower cell currents may go positive to near
the Vdd level. This would not cause any disturbance on the
addressed segment. If however the read bit line starts out at
Vref + 600mV and the clamp transistor turns off while the read

15   bit line is still more than 80mV above the reference, the write
· bit line would go negative.

       Only if the clamp is turned off very early could the
write bit line go below the word line voltage, thus turning on
the write transistor and write a zero in a cell that was a one.

20   This case is very remote as the word line voltage would by
design not be turned on that early.

       A more serious problem occurs in reading and writing
data in a segmented bit line configuration. The disturbance
introduced by turning off the clamp voltage would remain on the

25   non-selected segments. The current to the write bit line
caused by an 80mv differential voltage on the read bit line is
approximately 12µA, with the inverter dimension as discussed
above. The bit lines would be displaced at a rate of 120mV/ns
and thus completely override the contribution from the selected

30   segment except for rather high cell currents. A cell current
of 1µA will change the read bit line voltage at a rate of
10mv/ns. The displacement on the write bit line after T nS
follows the equation Vdispl = Icell(µA)*10*TSquared/2mV. A
cell current of 10µa would after 2ns have displaced the read

35   bit to +120mV and the write bit line to -120mv, while at the
same time the bit lines on the non-selected segments would have
moved 240mV positive. A simulation has shown that a cell
current of at least 40µA would be required to operate an

8-segment memory, assuming bit line capacitances of 100 fF. By
using different clamping times for selected and non selected
segments, the problem is highly reduced. If the clamping on
the non selected segment is terminated just before the signal
5   coupling the segments together is turned on, the voltage
between the non-selected segments would be approximately 100
mV. Allowing the bit lines in the selected segment to be
displaced more than k*100 mV before coupling, where k is the
number of segments, would yield more reliable operation. Some
10 of the amplifiers shown may use a second inverter in each
segment which can be selectively engaged. Allowing the second
inverter to contribute to the amplification in the selected
segment before the coupling satisfies the above-stated
requirement.

15   It was assumed above that the power to the amplifiers
was selectively turned on and off during the memory cycles.
The method of clamping to the switching point of the inverters
used in some amplifier configurations, however, makes it
natural to operate the amplifiers under full power all the
20 time. This means that in standby mode, the bit lines will
always be at or close to the rail voltages. The clamping will,
when used in connection with an inverter, bring the bit lines
to the inverter switching point against the transistor that is
in the holding mode. The final adjustment to the switching
25 point level will therefore take somewhat longer than if the bit
lines start out already close to the switching point as
described before.

            Fig. 4A illustrates an amplifier 400 intended for
segmented bit lines. The basic inverting amplifier consists of
30 transistors 402 and 404 that make up the first inverter, while
the second inverter with transistors 406 and 408 may be added
for additional amplification later in the memory cycle.
Transistor 410, when turned on by the signal VINV2 will connect
the output of the second inverter to the input of the first.
35 The clamping is done by turning on the clamping transistor 412,
thus shorting the bit lines BLR and BLW and adjusting to the
switching point level of the first inverter. The selected word
lines on the selected segments are brought to the reading level

at essentially the same time as VCLAMP goes low. The read bit
line BLR at this time has a predictable offset error as
discussed above, due to incomplete adjustment to the switching
point during clamping and the capacitive coupling from VCLAMP,
5   when it goes negative. With only the first inverter connected
(VINV2 low), the cell current will both overcome the offset and
displace the bit lines sufficiently for a correct refresh
level. If limited to only the first inverter and no coupling
between segments, this would be a relatively slow operation of
10 high power consumption.

In the refresh mode, as discussed above, the segments
remain separated and refresh cells on their selected word
lines. To speed up the refresh cycle, the second inverter is
connected at the time when the bit lines are separated on the
15 order of 100 mV. The amplifier now acts as a conventional
latching amplifier. When the memory is used in read or write
mode, the offset on non-selected segment must still be
considered. At the time of connecting the segments, the
signals on the non-selected segments must be an order of
20 magnitude smaller than the signal on the selected segment. As
shown in the timing diagrams of Fig. 4B, a different approach
is used that guarantees a suitable ratio between the signals.
For the selected segment the clamping takes place first in the
cycle, allowing the signal to grow, driven by the first
25 inverter. Then the second inverter is connected and the higher
amplification proceeds for some time until the segments are
interconnected by the signal VCOUPL. A temporary drop in the
signal is followed by the mutual amplification from the cross-
coupled segment inverters. Initially only the selected segment
30 has its second inverter contributing. The non-selected
segments as shown in the lower timing diagram all have a
delayed clamp signal that is turned off just before VCOUPL is
turned on. Therefore, the only signal on the bit lines of the
non-selected segments is the offset signal caused by the
35 turnoff of the clamp signal. Once the coupling has been
established, the second inverters on the non-selected segments
are turned on, introducing additional speed to the read
operation. The bottom segments in the selected columns may use

the early clamp signal in write mode, permitting the full read cycle to apply the write signal, so that the signal at the bottom segment at the time of coupling is sufficiently larger than the signal on a selected segment other than the bottom
5 segment, to guarantee correct writing.

Fig. 4C shows amplifier 420 that is a variation of the two inverter amplifier. In amplifier 420, the clamping transistors include both N transistor 412 and P transistor 422, and coupling transistors include both N transistor 410 and P
10 transistor 424. By optimizing the design in this way, most of the clamping disturbance can be eliminated.

It has been mentioned that different timing is required for selected and non-selected segments. Figures 5A and 5B show the logic required for each segment for controlling
15 an associated amplifier. All segments have the same logic shown in Fig. 5A, except the bottom segment of a bit line, shown in Fig. 5B, that is used for transferring data in and out of the bit line. A memory cycle may be initiated by a separately timed refresh request or an interrupting Block
20 Enable or Chip Enable signal. Word Line Decoder block 502 and Segment Decoder block 504 decode the incoming function and address code to select a given segment in the memory and a given operation (a read or a write) and initiate a timing sequence. The appropriate timing signals are generated and
25 applied to the memory block. The signals are routed to all segment controllers and are used slightly differently for selected and non-selected segments. In the refresh operation, which is requested by a free running clock, in one mode all segments are considered to be selected. The applied timing
30 signals in the figures are Early Clamp Time, Late Clamp Time, Selected Segment Inverter on Time (SelInvTime), Non-selected Segment Inverter on Time (NonSelInvTime), Word Line on Time (WLonTime) and Dummy Current on Time (DummyOnTime). The abbreviations for these signals are shown in parentheses where
35 appropriate. A ramp signal VRAMP is also applied, which may be common for the block or be generated in each segment, with the timing controlled by common signals not shown on the diagram. The logic for the bottom segment shown in Fig. 5B operates

differently than the other segments during the write operation. The enabling of word line and dummy current is the same for all segments, but the clamp signal and the connection of the second inverter are always early for the bottom segment in the write
5 mode. This ensures that the written data gives a dominating signal to the connected segments. In write mode, input data is coupled to BLR of the bottom segment; in read mode, output data is coupled to BLW of the bottom segment.

Fig. 6A illustrates a memory 600 having two segments
10 602 and 604 where only one inverter 606 and 608, respectively, in each segment operates under fixed power. Read bit line BLR from segment 602 is coupled to write bit line BLW of the next segment 604 through pass transistors 610 and 612 when VCOUPL is high. Fig. 6B shows the relative timing between selected and
15 non-selected segments. In refresh mode, in a preferred embodiment, all segments are considered selected and the coupling signal remains low during the cycle. Alternatively, the segments may cooperate in pairs, resulting in the selection of every other segment. Fig. 6C shows how the selection of
20 segment for refresh is controlled by the odd/even bit. This decoding may be common for all segments or local as shown. The coupling signal is divided in two. One signal is used only in the refresh mode, while the other is used in read and write modes. The advantage in pairing the segments in refresh mode
25 is that the time and power consumption is reduced due to the higher amplification for the two cooperating inverters. The structure also makes it possible to pre-amplify the signal at reading and writing, before interconnecting all segments. The coupling signal interconnecting the two segments in the pair is
30 made subject to early selection if one of the segments is selected. This increases the amplification of the pair of signals. When a dominating signal amplitude has been reached, all remaining coupling signals are activated. For a preferred margin, clamping is performed at three different times. The
35 earliest clamp time is for the selected segment in a segment pair, the next clamp time is for the non-selected segment in the segment pair and the third time is for all remaining segments.

Conventional DRAMs and Segmented Bit Lines

Conventional DRAMs also gain in speed and margins by using the segmented bit line approach. The shorter bit lines used in conventional DRAMs have a lower capacitance that gives

5   stronger signal for given cell capacitor sizes. This approach actually permits the use of smaller cell capacitors and higher refresh frequencies, as discussed for the new cell type above. The same methods for minimizing the disturbance caused by the clamping operation apply to conventional DRAMs. There is,

10 however, a slight difference in the amplifier design. The cells in a folded bit line configuration may be alternately connected to one or the other of the two bit lines. All word lines with even addresses may, for instance, be selecting cells on one bit line, while all word lines with odd addresses would

15 select cells on the other bit line. In the simplest amplifier discussed earlier having only one inverter, the coupling between segments must complete the amplifier so that the inverter output of one segment connects to the input of the inverter in the next segment. Each segment must therefore have

20 two inverters, each selectively used according to which bit line is addressed.

Fig. 7A shows memory cell 700 including an amplifier where while reading cell 701 power is first turned on to inverter 702, whose input is coupled to read bit line BLR

25 holding the cell to be addressed by the word line. The bit lines are clamped while power is on, adjusting to the switching point of inverter 702. Due to capacitive coupling when VCLAMP goes low, the displacement on read bit line BLR is compensated by VCOMPODD going positive, as shown in Fig. 7B. The second

30 inverter 704 in a selected segment may be powered up before the segments are coupled together by the VCOUPL signal. For reading and writing operations, this gives a large enough signal to overcome the combined error signals from the non-selected segments as discussed above. The non-selected

35 segments will typically have the bit lines clamped late on the primary inverter 702 so that at time of coupling a very small signal has developed. The secondary inverter in non-selected segments is powered up only after the segments have been

connected. Only in the refresh mode are both inverters 702 and 704 used the same way in all segments. The same principles regarding the inconvenience of turning power on and off applies to conventional DRAMs, so the same approach discussed for Figs.

5   4A and 4B above can also be used here. The only difference is that, when used in connection with a fixed Vdd supply, the order of connection of the inverter amplifiers is determined on which bit line holds the addressed cell.

Fig. 7C shows a memory 710 including this type of

10   amplifier 712, with capacitive compensation of the disturbance at the termination of the clamping. The first inverter 714, with its input connected to the first bit line BL1 has its output connected to the second bit line BL2 when VINVEVEN goes high, while the second inverter 716 connects its output to the

15   first bit line BL1 at the time when VINVODD goes high. If, for example, the input of inverter 714 is coupled to the bit line of the selected cell, then inverter 714 is activated at the start of the memory cycle and is used in the clamping operation, which adjusts the bit line voltage to become close

20   to the switching point of the inverter. The second inverter 716 is then activated late in the cycle, when the output from the first inverter 714 is much larger than any offset error due to device variation between the first and the second inverter. For selected segments, the second inverter 716 is activated

25   some time before the signal VCOUPL is applied. This signal connects the segment to the neighboring segments for read and write operations only. In the refresh mode, all segments are selected and the read restore operation is completed with help of both inverters. The VCOUPL signal remains low during the

30   refresh operation. In the read and write mode, as described above, the second inverter in the non-selected segments is turned on after VCOUPL has connected the segments. Some time is allowed to permit the first inverters in all the non-selected segments, in cooperation with the two inverters in

35   a selected segment, to amplify the temporarily reduced signal enough to dominate over the coupling transient when the second inverters are activated. As indicated in Fig. 7C, the succeeding segment mirrors the previous segment, so that an

even addressed cell is located on the right bit line if the even cell on the previous segment was located on the left. Likewise the inverter direction is reversed in the succeeding segment (not shown in the figure). The timing relations

5   between the signals for selected and non selected segments are shown in Fig. 7D.

Fig. 7E shows the segment control logic for memories 700 and 710 discussed above. The segment control logic of Fig. 7E is similar to that shown in Figs. 5A, 5B and 6C above,

10 except for the separate control of odd and even inverter and clamp compensation.

Two Tier Memory

The DRAM of the present invention by its nature makes

15 it possible to design a memory where the content of the memory can be read without any interference from ongoing writing, reading and refreshing. The memory is structured in "two tiers," meaning that in addition to the first tier, a regular DRAM of the present invention with normal addressing, reading,

20 writing and refreshing, a second tier with separate addressing is used to read the contents of the cells in the first tier.

Fig. 8A shows a two tier memory 800 with first tier memory cell 801. The DRAM cell, which includes transistors 802, 804 and 806 with memory capacitor C1, is addressed by its

25 word line WLi and is treated as part of an independent memory in the reading, writing and refreshing operations. As discussed earlier, one of the features of this DRAM is that the stored data in the memory capacitor C1 is disturbed very little during the operation of the memory. As a matter of fact, the

30 high state need not be disturbed at all, while the disturbance in the low state (the voltage on C1 close to zero) can be limited to around Vth. For higher operating speeds a slightly higher disturbance is acceptable. The first tier memory 801 can have any of the features discussed above, including being

35 of the segmented bit line type.

In the second tier memory cell 810, transistor 812 is used to sense the state of C1 and transistor 814 acts as the read transistor. The second tier cell 810 is addressed by

WLTi, by applying a high signal thereon. This connects second tier cell 810 to the bit line BLT, which in turn connects to a sense amplifier 816. Amplifier 816 may, merely by way of example, be a single-ended amplifier, as the signal amplitude

5   is such that folded bit lines are not necessary for common mode noise suppression. However, folded bit line configurations 820 and 830, such as those shown in Figs. 8B and 8C, respectively, are also possible. Fig. 8D shows a pair of memory cells 842 and 844, each having first and the second tier cells, 846, 848

10 and 850, 852, respectively. The first and second tier cells share a common ground bus GND. The connections to the bit lines are common for neighboring primary cells. The second tier cells have separate connections to BLT, but using three level metal and different layouts, common connection points can

15 be used.

It should be noted that the sense transistor in all the previous figures has the source connected to ground as in most cells discussed above. The reversed cell, where the select transistor has the source connected to ground and the

20 sense transistor connected between the select transistor and the BLT, can also be used. This configuration is shown in Fig. 8E. As a rule, this cell requires a larger storage capacitor C1 than the cell with both sense transistors connected to ground. The gate capacitance of the sense

25 transistor in the second tier cell in read mode will couple a disturbance to C1, which may cause an error in the memory functions of the first tier cell. This is contrary to the normal (non-reversed) cell, where the major part of C1 actually is the sum of the gate capacitances of transistors 802 and 812

30 (Fig. 8A).

Again, the amplifier for the second tier may be of the folded bit line type, as indicated in figures 8B and 8C. This amplifier would use the same techniques discussed above in conjunction with memories having only one tier. Segmented bit

35 lines may be used in this configuration as well.

A single-ended amplifier 850 with a data out latch 852 is indicated in Fig. 8E, and a corresponding timing diagram is shown in Fig. 8F. The signal VBIAS is held at an

intermediate level concurrently with VWLT. The current from P
channel transistor 854 is approximately half of the sum of the
cell currents for a zero cell and a one cell. Limiting the
negative swing of VBIAS makes P transistor 854 act more like a
5 constant current source than if the swing is larger. The
circuit can be made to operate with somewhat lower margins if
designed so that VBIAS goes to ground when activated. The D
input on the flip flop is designed to have its switching point
close to Vdd/2, and as long as the bit line voltage in the two
10 states deviates a minimum amount from Vdd/2, the operational
margins are satisfied. The actual moment of reading may in the
worst case coincide with the reading of the first tier memory.
At this moment C1 may have a disturbed Zero level. The
amplifier design must take this into account and also plan for
15 changes in the charge levels on C1 between refresh due to
leakage. In the amplifier indicated in Figs. 8E and 8F, no
clamping of BLT to a neutral midpoint is made. Adding this
feature reduces the access time of the second tier memory 810,
but would not reduce the power consumption. The folded bit
20 line approach has both speed and power advantages, but the gain
is highly dependent on the bit pattern in the read data.

Denser layouts using a two-tiered memory may be
achieved by using non-dedicated bit lines, which means that a
given bit line can be used as a read bit line or a write bit
25 line depending on if the address is odd or even. In this type
of arrangement, the bit line capacitances remain equal.
Fig. 8G shows part of such a memory using the DRAM cell of the
present invention expanded to include a sense transistor and a
select transistor for the second tier. In the configuration
30 shown, each column of basic DRAM cells is combined with two
folded bit lines for read out in the second tier. Read
operations from this memory configuration are very fast. The
bit lines are clamped to a mid level, the word lines swing to
Vdd when reading, the amplifier is disconnected from the bit
35 lines at an optimum time when the bit lines have just separated
from the clamp level, and the reclamping is completed while
data is read out from the amplifier. A preliminary layout for

this circuit is shown in Fig. 8H.   Other configurations may of course yield a denser layout.

Process for Reduced Area and High Speed

5   One of the features of the memory of this type is that the conventional CMOS process with three metal layers can be used for production of the memory.   For dedicated memory chips, the cell area can be reduced substantially if the fused and thinned wafer approach is taken, as also described in U.S.

10   patent number 5,396,452.   In logic circuits where memory and logic is mixed, even larger benefits can be achieved.   The approach would be to prepare a wafer that, later in the process, would strictly follow a standard fabrication process. The added cost in the wafer preparation would be well

15   compensated by the savings in area and the gain in speed.

Fig. 9A illustrates steps in the wafer preparation as follows:

1.   A support wafer 902 of low resistivity is used. The areas 904 where capacitors are to be located are etched

20   down.   A thin dielectric layer 906 is deposited or grown.

2.   Polysilicon layer 908 is deposited over thin oxide layer 906.

3.   The wafer is polished, so that the support wafer 902 and the polysilicon 908 have a common smooth surface.

25   4.   Another thin oxide layer 910 is grown on the polished surface.

5.   The epi layer 912 of a second wafer faces the top of the support wafer and is fused thereto.   The assembly is then processed as described above so that only the epi layer

30   remains on top of the support wafer.

6.   Contact holes 914 are etched and plugs 916 are deposited to establish contact points to the embedded capacitor 918 and to the substrate.

7.   If other means have not been used in the step

35   above to isolate the plugs from the surrounding silicon, trenches 920 surrounding the plugs are defined and etched.   In the same step, the individual transistors in the memory cell and in the rest of the logic are also separated from each other

by trenches 920. The trenches are preferably back-filled with a form of glass.

8. P-channel transistors are exposed to N-type implant and are heat treated to a sufficient depth. The

5 surface is next prepared for the gate oxide. If the capacitor is to be connected to the gate of a transistor, the gate oxide 922 is removed on top of the contact plug for the capacitor. To more closely follow a standard fabrication process, the capacitor connection can wait for the metal 1 step to connect

10 the gate and the capacitor.

The wafer is now ready to be run in a known three layer metal process. The LOCOS isolation process step need not be completed, as the area under the poly contact can be isolated by a trench in step 7 above. This results in a planar

15 metal 1 layer.

Multi-layer Memory Cell Controlling Logic

Programmable logic devices such as Field Programmable Gate Arrays (FPGA) have devices that must be controlled with

20 relatively large capacitors. The pass transistors transferring signals from one data bus to another must be large enough for minimum voltage drop in passing the signals. The controlling capacitor must be significantly larger than the gate capacitance of the pass transistor, so that the voltage on the

25 controlling capacitor is relatively constant during the signal transition. Using the multi-layer approach described above lends itself ideally to this purpose. The total area under the pass transistor and under the memory cell transistors can be used for the controlling capacitor.

30 Fig. 9B shows one DRAM cell 930 controlling a pass transistor N4. The cell 930 includes transistors N1, N2, N3 and a large capacitor C1 located under both the cell and the pass transistor. The source of N1 is connected to the gates of N2 and N4 through first metal layer M1, with a silicon island

35 used to support a feed through to the underlying capacitor. (The vertical connection is indicated on the left side of the figure.) The combined structure has three vertical busses, which may be in the second metal layer M2, with two busses

reserved for the memory bit lines and one reserved for a
vertical interconnect bus for data. The word line and local
connections use layer M1, while ground and a horizontal data
bus use the top metal layer M3. There are several options to
5   make the connections between the different layers, including
feed-throughs in the back-filled areas between transistors.
The epi layer remaining after thinning the top wafer is usually
thicker than a normal implanted source-drain diffusion. The
transistors therefore have a remaining substrate portion.
10   Provisions have therefore been made so that this substrate
layer may be connected to the underlying support wafer which is
at ground level for the N-channel transistors. The substrate
layer may be connected to the underlying support wafer which
may be at a ground level (Fig. 9C). The support wafer does
15   have N wells to support the connection of the substrate of the
P transistors for this reason.

        The same procedure discussed in conjunction with
Figs. 9A and 9B can also be used for the memory cell shown in
Fig. 9C. Adding a recessed capacitor 936 increases the storage
20   time of the cell, which reduces the refresh frequency. The
gate capacitance of the sense transistor is generally
sufficient for reliable operation. With all three cell
transistors in the top layer, no additional area is gained by
adding a recessed capacitor.

25   Fig. 9D shows in greater detail a compact cell
utilizing a recessed transistor and a double-sided capacitor.
The thin transistors in this figure are not fully depleted,
which means that the transistor substrate must be tied to
ground. It should also be noted that the transistor substrate
30   is one side of a double sided capacitor and for that reason
must have a relatively high conductivity. The surface of the
wafer to be fused should therefore have a P+ top layer. With
the transistor substrate grounded, the source and drain
junction capacitances are as high as in the standard processes,
35   while the area efficiency is increased as the devices are
separated by trenches cut through the top silicon layer. The
top transistor layer in this case is assumed to be formed by
defining an etch stop by an epi layer with a much different

impurity concentration and a different type. The etch rate in the bulk is therefore much higher with the right choice of etchant. The remaining layer is therefore too thick to permit full depletion by the junction implants while still maintaining

5   the small device dimensions.

An alternative approach has been taken in Fig. 9E. Here the top wafer used for fusion is of a SIMOX type. The top of this wafer has a thin silicon layer separated from the bulk by an implanted layer of oxygen. Such a wafer after adding a

10 relatively thick oxide layer, can be fused and thinned down to the buried oxide layer with high accuracy. The transistors formed in this thin layer will have fully depleted junctions and small junction capacitances. They will therefore be much faster than the transistors in Fig. 9D. The recessed capacitor

15 will however have a thicker oxide layer on top, resulting in a value close to half of the fully double sided capacitor in Fig. 9D.

Fig. 9E is a supplement to Fig. 9D, where marker line 942 denotes the location of the capacitor and marker line 944

20 the location of the junctions in the support wafer. In the cells of figures 9D and 9E, the bit line connections are not side by side. This permits the layout of a narrower cell than if the connection are side by side, which must be the case for single layer cells.

25 Fig. 9F shows another construction for a capacitor in combination with a pass transistor in a programmable logic device. In Fig. 9B, the capacitor was formed underneath the pass transistor. Here, the capacitor is instead formed above the pass transistor. In Fig. 9F, a flat capacitor covers most

30 of the pass transistor area and typically comprises two polysilicon layers separated by a thin dielectric layer. One layer is coupled to a fixed potential, for example ground or Vdd, while the other layer is coupled to the control node from the DRAM circuit. The gate of the pass transistor is likewise

35 coupled to the control node of the DRAM circuit.

Multi-layer Designs using Epitaxy and Simox

Figs. 10A-10D show alternate methods of submerging capacitors and transistors below a thin epi layer added directly on top of a supporting wafer.

5 Fig. 10A shows the steps for making a capacitor with both electrodes accessible. The sequence of processing is as follows:

a) A recess 1002 is etched in the support wafer 1004, which is assumed to be P-type.

10 b) N+ material 1006 is implanted but not necessarily heat treated at this time. The purpose is to form an N-channel junction, which will isolate the bottom electrode of the capacitor from the support wafer when operating at levels positive in relation to the support wafer.

15 c) A thin oxide 1008 is grown which will become the capacitor dielectric layer. Polysilicon layer 1010 is then deposited until the recess is filled. This can be a blanket deposit (as shown) or a local deposit.

d) The wafer is polished so that the remaining
20 surface 1012 is a continuous flat surface, but with no possibility to contact the buried N+ junction.

e) An area covering the capacitor and the intended junction contact regions is defined and another N+ implant 1006 is made.

25 f) An epitaxial layer 1014 is grown, which may require two steps to obtain a usable layer on top of the polysilicon. In the first step the layer is grown vertically and should give single crystal areas, at least where the starting surface is single crystal. In a second step some form
30 of zone crystallization may be used to include the areas above the poly into the single crystal.

g) In this step O_2 is implanted to a depth equal to the deposited epi layer. After heat treatment, an insulating layer of SiO_2 is formed. The remaining layer of the epi layer
35 is then used for MOS devices of the type used in Silicon on Insulator technology. Connections from the device layer to the capacitor electrodes uses known methods.

Fig. 10B illustrates how a transistor such as N1 in the memory cell can be submerged into the support wafer as follows:

a)   A recess 1016 is etched in the P- wafer, poly is
5 deposited and the wafer is polished.

b)   The transistor area is defined and N+ 1018 is implanted.

c)   The epi layer 1020 is grown as discussed above.

d) O_2 is implanted. When the wafer is heat
10 treated, SiO_2 is formed, leaving a device layer insulated from the submerged transistor. The transistor channel goes from the edge of the source implant to the edge of the drain implant. The channel will go in a vertical direction close to source and drain, if the depth of the diffusions are less than the depth
15 of the original recess.

Since there is some uncertainty about the silicon quality in the epi layer above the submerged poly layer, it is desirable to consider the approach illustrated in Figs. 10C and 10D, where the capacitor electrode is made from silicon. The
20 process for Fig. 10C is as follows:

a)   A deep N+ diffusion 1022 defining the area of the capacitor and its connection regions is made in a P type support wafer.

b)   To form a capacitor, dielectric layer 1024,
25 which may be, for example, N_4 or O_2, is implanted in a defined area.

c)   To isolate the capacitor electrode from the rest of the wafer, a trench 1026 is cut around its edges. The trench is refilled with SiO_2 using known techniques.
30   d)   Fields of epi layers 1028 are next grown on all silicon areas.

e)   Oxygen or nitrogen is now implanted to the depth of the epi layer and the completed wafer is heat treated. The top layer will be used for devices and conventional steps are
35 taken to contact the devices to the capacitor electrodes.

The method in Fig. 10D shows the process steps used to form a transistor with a large gate capacitor. As in Fig. 10C, the capacitor electrode (gate) is in crystalline

silicon, isolated from the rest of the support wafer by
implanting N_4 or O_2 and by trench isolation. It should be noted
that a capacitor can be made using the same configuration, but
by excluding the drain diffusion and connection.

5   Fig. 10E shows a conservative layout of cells using
submerged transistors.

Shared Bit Lines and Amplifiers

          In the single layer cells the bit line connections
10   are as mentioned above, side by side. Figs. 11A-C show an
alternate method by which bit lines are shared. This is
possible if different word lines access successive cells and
three metal interconnect layers are utilized. In Fig. 11A,
each bit line can be a read or a write bit line depending on
15   whether an odd or an even word line is addressed. For an odd
word line (WL1), the left-most bit line serves as a write bit
line and is therefore connected to the write bit line input of
the amplifier (control line "odd wlns"). The middle bit line
serves as a read bit line and is connected to the read bit line
20   input of the same amplifier. For an even word line (WL2), on
the other hand, the middle bit line serves as a write bit line
and is connected to the write input of the amplifier and the
third bit line serves as a read bit line and is connected to
the read line input of the amplifier. At the same time, the
25   left most bit line also acts as a read bit line and is
connected to the read bit line input of the amplifier on the
left (only the read line input is shown). In the figure, the
word lines are routed across the array in Metal 2. The Metal 1
is then used to reach the Poly gates, with the gate connection
30   located in the area of local oxide between the write and read
transistors. In the serially segmented bit line configuration,
the output from each segment amplifier connects to the bit
lines in the next segment via a pair of pass transistors as
shown for the dedicated bit line configurations discussed
35   earlier. In Fig. 11B the bit lines are dedicated for read line
or write line, but still with the feature of sharing. Here,
the second bit line is used as a write bit line both for the
cell on its left and the cell to its right. This bit line

always connects to write bit line input on the same amplifier, while the read bit line input to the amplifier is connected to the first or the third bit line depending on if an even or an odd word line is addressed.

5   The cells in Figs. 11A and 11B were assumed to be in a single level process. Further area savings will be achieved if the multi layer process is used also for the shared bit line configuration, as shown in Fig. 11C.

10 Refresh Read and Write in a Segmented Configuration

Fig. 12A indicates a portion of representative column 1202 in a segmented memory 1200 that may be selected for reading and writing. In the figure it is assumed that the first data bit in a data word of N bits can selectively be

15 connected to one of the first eight columns in the memory. Three column address bits, B0, B1 and B2, and their complements are used for addressing. Data can be read from or written to either bit line in the last segment 1204 (SEGMENT 0) in column 1202, but there is a small advantage in writing to the read bit

20 line and reading from the write bit line. The coupling is therefore done accordingly, controlled by the signals COPLIOREAD AND COPLIOWRITE.

The timing of the clamping and coupling signals for the segments are optimized for speed and performance. In read

25 mode, it is preferred that an addressed segment be allowed to amplify its signal before any additional segments are connected. In one embodiment, a first increment of time is set aside for the addressed segment, such as, for example, segment 1204, to establish a signal. At the end of this first time

30 increment, the second segment in the pair, which would be segment 1206 in this example, is coupled to first segment 1204 by a signal COPL. At the end of a second time increment, the remaining segments are interconnected.

However, if an inverting amplifier is used in the

35 memory segments, the segment pairs may be allowed to be connected at the outset of a read operation. Furthermore, if the dimensions of coupling transistors 1208 are chosen correctly, all segments may also be connected at the outset of

a read operation. This results in simplified logic and in a higher operating speed. In this simplified scheme, all segments are clamped at the same time at the beginning of the cycle followed immediately by an active coupling signal COPL

5 that connects the segments. With inverting amplifiers, memory refresh occurs in segment pairs; all segment pairs are refreshed at the same time. The signal COPL connects the two segments in the pair. This signal may also be applied after a delay that giving the selected segment signal time to establish

10 itself, compensating for any disturbance introduced by the clamp signal. On the other hand, coupling the two segments in a pair with COPL just after the CLMP signal is active is equally reliable, as both segments are disturbed equally. The fact that read bit line BLR in one segment connects to write

15 bit line BLW in the next segment results in nearly full compensation, without using a compensating signal as discussed earlier. The COPLR signal, which is used to connect one segment pair to an adjacent segment pair, is not applied in the refresh operation. For reading and writing, both COPL and

20 COPLR are applied after CLMP at the beginning of a cycle. A write operation must apply a signal to segment 1204 which is larger than the signal in the addressed cell if the cell happens to hold data of opposite value. If the topmost segment in column 1202 is being written to, the written signal would

25 have to travel through all the segments in column 1202 to reach the segment at the top of the column. At the same time, a signal of the opposite value may be travelling from the top segment in column 1202 toward the bottom segment 1204. At some point, probably near the middle of column 1202, the two

30 opposite-going signals would meet, and the difference between the two signals would then proceed in both directions. If the written signal dominates, the addressed segment at the top of column 1202 will have the correct data written. The larger the written signal, the faster the write bit line in the addressed

35 segment will reach a safe write level.

Only selected columns are addressed for the purposes of reading and writing data to the memory segments in the selected column. In both operations, the stored data in the

non-selected columns is not modified. All columns are exposed
to the same "horizontal" signals such as COPL, COPLR, CLMP, WL
and DUM (defining the dummy current level).

Figs. 12B through 12E are simulation outputs. A
5   memory array was simulated that included 256 columns with 16
segments per column arranged in 8 segment pairs, where each
segment had 64 memory cells. A 0.5 micron industrial process
at normal operating conditions was used. A 15ns cycle time was
simulated, which provides sufficient margins for reliable
10 operation.

In Fig. 12B, the simulation shows the results when a
memory cell storing zero volts on the cell capacitor is
addressed by a word line in the top segment. The write signal
is applied to the bottom segment, with a polarity such that the
15 stored zero will become a one. After the clamp and activation
of the coupling signal, the dummy current dominates over the
cell current and the read bit line for the addressed segment
(BLR15) starts going positive and write bit line for the
addressed segment (BLW15) goes negative. At the same time BLW0
20 is pulled low, causing BLR0 to go high, thereby initiating
writing of a one in the addressed cell. After approximately
9ns the written signal reaches the top segment, causing BLR15
and BLW15 to change direction of transition. BLW15 reaches the
full 5 V level at the end of the cycle and the cell is charged
25 to about 3.3 V, which is well above the level of 2.5 V that is
generally required to store a one in the cell capacitor.

In Fig. 12C, the simulation result is shown for the
operation of writing a zero over a stored one in the cell
capacitor in the top segment. The cell has a stored voltage
30 level of 3.5 V, which is the highest stored one voltage level
that can be written without boosting the word line level. This
gives the highest possible gain contribution from the stored
cell, which is acting against the written data input applied to
the bottom segment. The direction of BLW15 starts turning
35 around after 8ns and reaches 0.3 V at the end of the cycle.
The cell voltage tracks BLW15 after 11 ns and reaches 1.5 V
after 13ns. A cell voltage of 1.5 V is marginally acceptable

for a stored zero level, so the 15 ns cycle time provides
sufficient margin at normal operating conditions.

The stored data in non-selected columns is not
disturbed, as illustrated by the simulation results shown in

5   Figs. 12D and 12E. In Fig. 12D, a stored zero at a level of 1
V is originally present in the addressed cell in the top
segment. Bitlines BLR15 and BLW15 are the first to move due to
the dominance of the dummy current to the selected segment over
the current from the selected cell. BLW15 falls below 1 V

10 after 9 ns and reaches close to 0 V at the end of the cycle.
There is less than a 2 ns delay between BLW15 and BLR0 and an
actual crossover between BLR15 and BLW0, caused by the fact
that the coupling signal amplitude is limited to 5 V in the
simulation. The coupling transistors are not conducting higher

15 than one threshold voltage below 5 V. Due to a high body
effect of the transistor, the threshold voltage Vth at 3.5 V is
actually 1.5 V. As the same control signals are used for both
reading and writing on non-selected columns, Fig. 12D also
illustrates the read operation. The differential signal

20 between BLW0 and BLR0 represents the stored data. Fig. 12E
shows the simulated result when a stored one at 2.5 V is read.
The differential signal between BLW0 and BLR0 also here
represents the stored data.

It is apparent that the read data can be read out

25 long before the end of the full memory cycle. This is a very
important feature of this type of memory. A memory read or
refresh cycle can be interrupted at any time in the cycle
without destroying the stored data. A read cycle can therefore
be terminated as soon as a detectable differential signal

30 between the bit lines in the bottom segment is transferred to
the output. The only disadvantage is that the read cell level
will not be refreshed in an interrupted cycle, but the level
will remain the same as at the start of the cycle . The
voltage level in the cell will therefore remain the same until

35 the next refresh cycle.

If an I/O latch 1210 is added to memory 1200 from
Fig. 12A, as indicated in Fig. 12F, the delta V in Figs. 12D
and 12E can be allowed to be as low as 100 mV. This would mean

that data could be transferred to the latch after only 5 ns. Additional time is naturally required to forward the data to the output, but a read access time of less than 8 ns is practical for this array configuration.

5   The number of cells per segment has an impact on speed, which makes it possible to reduce the cycle time for reading and refreshing if the number of cells per segment are reduced. The delay through the coupling transistors combined with a limited gain in the amplifiers makes it difficult to
10 write over previously stored data in the most remote segment without making special arrangements. Simulations run on a memory structure with 256 columns, 8 segment pairs (16 segments) and 32 cells per segment are shown in Figs. 12G-12L. If a cell in the most remote segment is addressed for writing,
15 a read cycle to this segment will progress quite far before a write signal of opposite polarity reaches the addressed segment. On the other hand, if the bitlines in the segments of the addressed pair are held clamped for an extended time, while awaiting the arrival of the write signal, then the change of
20 polarity between the bit lines will be faster. In Fig. 12G, the clamp signals CLMP and CLMPR for the selected pair are allowed to stay on 2ns longer than the clamp signals for all other segments. All segments are coupled together at the same time indicated by COPL becoming active. The write inputs are
25 applied to the bit lines in the bottom segment (BLRR0 and BLWR0). The write signal proceeds from segment to segment toward the top of the array. Before it reaches the addressed segment, its clamping is terminated and the write signal overcomes the small signal from the cell. At a time of 7.5 ns,
30 the previously stored zero has been changed to an acceptable one. After 10ns, a maximum level for a stored one is reached. Fig. 12H shows the signals for writing a zero over a stored one.

    It is important that the columns not selected by the
35 column decoder for writing still preserve the stored data in their addressed cells. Fig. 12I shows how a stored one is actually restored to its maximum value in the non-selected column. BLR and BLW in the top segment still have enough time

to reach their full value. There is however not enough time for the signal generated in the top segment to proceed down to the bottom segment and overcome a noise-generated signal therein. The polarity of the bit line signals in this case

5   does not matter as the output is not used in this mode. Fig. 12J shows how a stored zero is restored in a non-selected column during the write operation. In this case, the noise signal in the bottom segment happened to go in the same direction as the signal coming down the line from the addressed

10   segment. The noise source for the bottom segment will be discussed in connection with the read mode below.

Figs. 12K and 12L show simulations of the read mode. To make sure that a signal from the top segment reaches the bottom segment and is correctly presented to the I/O circuits

15   some improvements have been made compared to the sequence used for the 64 cells/segment sequence described in connection with Figs. 12B-12E. The negative-going clamp signal will, as discussed earlier, introduce a small negative displacement of both bit lines. In a disconnected segment this would cause the

20   write bit line to move positive with the read bit line remaining stationary. In a long chain of segments, write and read bit lines are alternately connected, so the inside of a long chain is more or less self compensated. The problem arises in the segments at the ends of the chain. For all

25   segments the coupling signals adds a positive compensation. The top segment is exposed to one half compensation from the coupling signal, while the internal segments each are exposed to full compensation. The bottom segment in read mode is coupled to the I/O circuit through some form of decoder or to

30   an output latch. In either case the coupling transistors controlled by the signal COPLIN connect to nodes with unknown potential and capacitance. It is therefore desirable to connect these nodes to the bit lines during clamping to equalize the potentials. The COPLIN signal is turned on

35   momentarily at the beginning of CLMP and turned on again at the same time as all other coupling signals. Using this technique, Figs. 12K and 12L show how the bottom segment bit lines BLRR0 and BLWR0 correctly reflect the stored data.

In the previous discussion, it was assumed that the segments operated in pairs, so it is reasonable to use a segment pair decoder 1220 shown in Fig. 12M. The low order bit B0 in the segment address is used to select one or the other

5   segment after the higher order bits select the pair (B1....Bn). The dummy current for instance is applied only to the selected segment, so each segment has its own dummy voltage generator. The dummy transistor in the selected segment actually mirrors the current flowing through the dummy voltage generator. The

10   current through the dummy voltage generator 1222 is primarily a function of the RAMP voltage and increases the dummy current proportional to how the cell current varies with the RAMP voltage. The two transistors 1224 and 1226 between the RAMP-controlled transistor 1228 and ground are larger and serve

15   mainly as switches. Transistor 1224 is used to define how long the dummy current is to stay on during the cycle and transistor 1226 is turned on only during memory activity and when the specific segment is selected. A three input NOR gate 1230 combines the selected pair signal with the ODD/EVEN selection

20   by B0 and its complement to select the segment and associated dummy voltage generator. The fourth N-channel transistor 1232 in the dummy voltage generator is very small and in standby barely keeps the P-channel mirror transistor 1234 conducting.

A number of global timing signals are connected to

25   all segment pairs. Specifically, DUMTM defines the active dummy current time, ECLTM defines early clamp time, LCLTM defines late clamp time, ECPLTM defines early coupling time, and LCPLTM defines late coupling time. The complement of the refresh command (REFRB) is also a global signal. REFRB (low

30   level) bypasses the segment pair decoder and applies an early clamp signal to both segments through signals CLMP and CLMPR. An early coupling signal is also applied to the top segment in each pair, connecting the two segments in the middle. The LCPLTM signal is not delivered in the refresh cycle, thus

35   keeping all pairs separated.

Control of Logic Cells with DRAMs

U.S. Patent Numbers 5,375,086 and 5,317,212, both hereby incorporated by reference, describe methods for controlling the function of logic by using the capacitors in
5   DRAMs as the controlling element. The DRAM of the present invention lends itself to control of logic with some advantages over the previously disclosed method.

The control capacitors that are also storage capacitors in the memory cells are typically larger than what
10   is required for memory functions. With the leakage currents in the dielectric of the capacitors much smaller than the leakage in the write transistor, successive refresh cycles can occur less frequently. As most of the power at refresh is consumed moving the bit lines, the refresh power is inherently low.
15   However the power consumption can be further reduced by using segmented bit lines, as the power consumption is inversely proportional to the number of segments.

The control capacitor Cc, which principally replaces an SRAM cell, must in its high state be at least at a level of
20   Vdd. The cell capacitor is typically connected to the gate of a large pass transistor, which has a gate capacitance of Cgp. Any signal Vsig passed by the pass transistor will be coupled to Cc through a capacitive voltage divider. The variation on the voltage on a programmed cell is described by
25   $Dvc = Vsig \times Cgp/(Cgp + Cc)$. The signal Vsig typically has an amplitude of Vdd. If Vdd is 5 V and $Cc = 4 \times Cgp$, then Dvc will be 1 V. The programming is asynchronous to any signals in the controlled system. If programming occurs when the passed signal is at Vdd and the programmed cell voltage is Vcellp,
30   then the cell voltage will be reduced by Dvc, when the passed signal has returned to ground. The gate to source/drain voltage will go from (Vcellp - Vdd) to (Vcellp - Dvc) during the negative transition. Conversely at a following transition the gate to source/drain voltage will go
35   from (Vcellp - Dvc) to (Vcellp - Vdd).

If the programming occurs when the passed signal is at ground level, then the cell voltage will increase by Dvc when the passed signal goes from 0 to Vdd. The gate to

source/drain voltage will go from Vcellp to (Vcellp +
Dvc - Vdd) during the positive transition. If the design goal
is to equal the performance of SRAM control, then Vcellp is at
a level of Vdd. The SRAM has a relatively high output

5   resistance, so in a negative transition the gate voltage will
drop and recover with a time constant of (Routput x Cgp). If
it is assumed that this slows the negative transition
approximately as much as the transition would be slowed by
reducing the gate to source voltage to Vdd - Dvc at the end of

10   the transition, then a relatively high value such as 2 V could
be chosen for Dvc. When the positive transition occurs
(assuming that the programming was done when the signal was at
Vdd level) the gate to source voltage would go from 3 V to 0 V,
while the SRAM-controlled gate to source voltage would go from

15   Vdd to a positive voltage caused by the same time constant of
(Routput x Cgp), but limited by the forward-biased diode from
the drain of the P-channel transistor to its N-well
(approximately 0.6 V).

If the programming had occurred at a low signal, both

20   transitions would have been faster than for the SRAM case, with
the gate to source voltages (assuming Vdd + 5 volt) 5 V and 7 V
in the two extremes of the transitions. To program Vcellp to
5 V, the supply voltage to the word line of the amplifier must
be approximately 6.6 V due to the high body effect on the short

25   channel transistors. If Dvc = 2, the cell voltage would peak
at 7 V. The pass transistor gate would however be exposed to a
maximum of 5 V, while the full 7 V would stress the gate of the
sense transistor and the extended storage capacitor. If 7 V is
the specified maximum voltage, the word line voltage would also

30   be permitted to go that high. The amplifier supply could then
be 5.4 V and a Dvc of 1.6 would be chosen. The cell
capacitance is determined by the relationship
Cc = (5 / 1.6 - 1) x Cgp = 2.125 x Cgp. A pass transistor may
have W = 20 μ and L = .5 μ, or a total gate area of 10 μsq.

35   The cell capacitor area would be 21.25 μsq or in the form of a
full transistor with W = 4.6 μ and L = 4.6 μ. The pass
transistor would occupy about 70 μsq, while the capacitor would
occupy 35 μsq. A capacitor with a source in the middle would

only occupy 25 μsq. It is apparent that Dvc could be reduced even further at relatively small area cost.

The actual voltages and capacitor sizes depend on process limits and design criteria.

5   The DRAM cell of the present invention has been described as using three N-channel transistors, but it should be understood that three P-channel transistors can naturally also be used, with associated circuits and signals modified to fit the reversed requirements.

10   The addressing of a memory used for control of logic need not be as fast as for a conventional memory. Data and addresses can be fed serially through shift registers in both vertical and horizontal directions. Recent requirements for higher speed of loading uses byte wide data shifting and for

15   more advanced applications random addressing for the writing of control data is used.

The Fig. 13A shows a decoder 1300 that may be used to address the word lines. The decoder is operated by N pairs of signals, both true and complement, selecting a desired word

20   line. The SEGMSEL signal may be global for non-segmented memories or may be derived from a combination of a global enabling signal and a segment decoder. The low order bit An+1 and its complement selects one segment in a pair to be fully decoded for read and write operations, while the higher order

25   bits are bypassed for the refresh operation. The input to the first inverter 1302 in the segment select circuit is assumed to have a weak pullup device normally. With the decoder activated, the input node is pulled low and the output of the first inverter 1302 enables all word line decoders in the

30   selected segment. However only the word line selected by A0 to An and their complements will track the global VRAMP signal. The output of second inverter 1304 goes low on the selected segment, thus allowing the selected word line to go high.

The active time for the cycle is superimposed on An+1

35   and its complement such that both are low between cycles. Then at the end of the cycle, the input to the first inverter 1302 will go high, as will the output of the second inverter 1304. Transistor 1306 will then pull down the selected word line and

hold it and all other word lines to ground. To reach a word line voltage of 7 V, VRAMP need not go higher than 7 V, but the address inputs and the output of the first inverter must reach one threshold higher, which means close to 9 volt. This type

5   of decoder 1300 would therefore be used in connection with a large ratio between the cell capacitance and the load capacitance (gate capacitance of the pass transistor), so that the operating voltage could be reduced.

Fig. 13B shows a decoder 1310 with a boosting feature

10 which allows most of the circuits to operate at normal Vdd levels. A WLENABLE and a complement WLRESET is applied globally or from a segment select circuit. In the enable mode, LIN on the decoded circuit is pulled low, setting the previous reset latch. When the LIN signal is low and the LOUT signal is

15 high, nodes DN2 and DN3 are elevated to a level that is one threshold voltage below Vdd (e.g., 3.5 volt). The timing diagrams in Fig. 13B show the levels of the different nodes. The RAMP signal starts out at a low voltage level of approximately 1.2 V, and WL is brought to that level by fully

20 turning on MNWL. Late in the cycle, RAMP moves from 1.2 V to 8 V, in this particular case. The capacitor MNC is much larger than the capacitance of node DN2, so DN2 will track the rise of the RAMP and the word line. DN2 will end up higher than the RAMP by one threshold voltage, so the word line will go as high

25 as the RAMP signal (e.g., 8 volt). DN2 at 9.6 V might expose the gate oxide of MN3 to the break down limit. Thus, MN2 acts as a barrier in that its gate is at 5 V, as is the gate of MN1. By returning the RAMP voltage to 1.2 volt before resetting the word line, the protection of the devices is maintained.

30 The voltage of a stored one with this conditions would be approximately 6.4 V, provided the amplifier supply voltage is also at 6.4 V. The maximum stress on gate oxides in the cell is 6.4 volt plus Dvc.

When controlling logic, the high voltage on the cell

35 is of most concern, even when an stored zero at a level of about 1 V may be acceptable. Another alternative is, therefore, to implement a decoder using P-channel devices only. Fig. 13C shows decoder 1320 that is the equivalent of Fig. 13B,

but with the polarities reversed. The common node for the
sense transistors and cell capacitors in the cells are
connected to the 6.5 V supply indicated in the figure. The low
RAMP level is -1.5 V in relation to 6.5 V or the 5 V Vdd

5 supply. To approach the Vss level in the controlled circuits,
it is desirable for the RAMP to go to -1.5 V, thus overcoming
the high threshold voltage due to the body effect of the
P-channel write transistor in the cell. The negative swing on
the RAMP signal could be limited to 0.5 volt below Vss. The

10 refreshed high voltage on the controlling capacitor will be at
6.5 V, but in the read phase of the refresh operation
the voltage will drop slightly. For cells with large
capacitors that are used for controlling pass transistors,
this voltage drop is very limited due to the long time constant

15 in the combination of the write transistor and the cell
capacitor. In other words, the write bit line will reach the
6.5 V level before the cell capacitor has been disturbed more
than a small fraction of 1 V. The refresh phase with the RAMP
and word line at its negative extreme is made long enough to

20 fully refresh both a high and a low cell level. The original
storing of control data may require extra long write times or
multiple write cycles for very large storage capacitors. All
cells can also be precharged to an intermediate level by
turning on all word lines, with the bit lines clamped to the

25 intermediate level. The writing will then only need to
displace the cell voltage slightly to indicate the stored data.
The refresh operations will later gradually charge the
capacitors to their full values.

30 <u>Switched Power Amplifier</u>
           In order to further reduce the power consumption in
the control memory, it is desirable to let the cell current or
the dummy current displace the read bit line as much as
possible before the amplifier is powered up. Fig. 13D shows a

35 memory 1330 where each pair of segments 1332 contains only one
common clamp circuit. The associated timing diagrams are shown
in Fig. 13E. Clamping to an intermediate clamp level is done
without power to the amplifier and further timed so that at

least the two segments in each pair are coupled together during
the clamping. All four bit lines will therefore be floating
close to the clamp level at the beginning of the memory cycle.
Applying the low ramp level to the word line starts the
5 displacement of the read bit line in the selected segment.
When the selected bit line has been displaced sufficiently for
good operating margins, the power is applied to the selected
segment. This is controlled by an early timing signal, EPERB,
which is negative true. When the inverting amplifier in the
10 selected segment has displaced its write bit line a certain
amount, power is applied to the non selected segment in the
pair (controlled by LPWR) and the two segments 1332 in the pair
are coupled together to form a full amplifier. After
increasing the ramp voltage and holding it at its high level
15 for a time, the cell is refreshed and the cycle is terminated
by turning off power, clamping, and coupling all segments
together. In the read and write modes, indicated as REFRB, all
segments are coupled together at the early coupling time
(ECPLTM). As described in connection with a regular memory, a
20 very large signal is applied to the bottom segment of a
selected column which dominates over the signal detected at the
selected segment.

Mixed Cell Types on Common Bit Lines

25   Typically large cell capacitors are required to
control the gate level on pass transistors. At the same time,
the bit line capacitances can be kept low by utilizing the
segmented bit line approach. The ratio of the cell capacitance
and the bit line capacitance will in such cases be high enough
30 to impose a large bit line displacement at reading a
conventional DRAM cell. In other circuits controlled by the
control memory, such as look up tables, the cell capacitor can
be very small and the preferred cell is of the new DRAM type.
It is possible to have both types of cells on common bit line
35 pairs, as indicated in memory 1340 shown in Fig. 13F. Here, an
approach with non-dedicated bit lines has been used, which
means that a given bit line can be used as a read bit line or a
write bit line depending on if the address is odd or even. In

this type of arrangement, the bit line capacitances remain equal. The dummy current is applied to the bit line that is active for reading a new type of cell and is applied to the bit line to which the selected conventional DRAM cell is connected.

5   The dummy charge at reading a conventional DRAM cell is the product of the dummy current and the dummy current on-time, which operates with the same margin as for the new DRAM cell of the present invention. The bit lines are clamped to an intermediate level (e.g., 3.5 V) at the start of the cycle,

10 which is higher than the first RAMP level of 1.2 V. A fully charged conventional cell capacitor (e.g., 5.6 V) will also be well above the word line voltage of 1.2 volt. The cell transistor is off and the dummy current increases the bit line voltage at the same rate as at reading a new type of cell.

15 If, on the other hand, the cell capacitor is discharged to a low level (e.g., 0 V), charge is drawn from the bit line, reducing the bit line voltage more than the dummy current is able to compensate before the amplifier dominates the displacement of the bit lines. With a normal RAMP signal, the

20 word line voltage is increased later in the cycle bringing the cell capacitor to its refreshed level, also for a stored one. In the word line, addressing is sequential without regard to what type of cell is used. In Fig. 13F, WLA0 and WLA1 indicate even and odd word lines addressing conventional cells, and WLB0

25 and WLB1 indicate even and odd word lines addressing the new type of cells. The dummy voltages DUM0 and DUM1 are active negative and turn on the dummy current on BL2 and BL1 respectively. Note that to write a one (high voltage on the capacitor) on an even cell of the new type, BL1 must be forced

30 high by the write circuit, while on the other hand BL2 must be forced high to write a one on an even cell of in a conventional DRAM. The programmer must therefore take this into account when preparing the control data. Only if cells of the same type are used for a given range of addresses an alternative

35 approach can be taken. In that case, using the figure as a basis for the discussion, even addresses would apply to conventional cells placed on BL1 and DUM1 would be activated for all even addresses calling this type of cells. This would

however place unnecessary restrictions on the cell distribution.

The timing for refreshing a stored zero and a stored one in a conventional cell is shown in Fig. 13G. Note for the
5 stored zero that the exchange of charge continues only until the cell capacitor has been charged to one threshold voltage below WLA0. This also points out the need for frequent refreshing as a cell charged by leakage current to close to one threshold below the initial word line voltage will cause a read
10 error. In reality, the large cell capacitors used in this configuration at the refresh frequency dictated by the new type of DRAM cells will be charged very little between cycles. The COUPL signal and the disturbance on the early bit line signals have not been shown in this figure as this is of interest only
15 in the initial write mode and in the read mode used for checking that the written data was correctly stored (mainly in production tests).

Disabling Segments Above Selected Segments
20 In both read and write operations, the segments above the selected segment need not be activated. This is done by forwarding the segment select signal to the segment above, as depicted in Fig. 14A. In each segment an OR is formed by the segment select signal and the selected signal from the segment
25 below this or signal is then forwarded to the segment above, where it is treated as the select signal from the segment below.

One disadvantage with this approach is the serial delay time to the last segments in the chain. It is desirable
30 to be able to interrupt the clamping signals in all segments before they occur. A parallel approach as indicated in Fig. 14B addresses this concern. In this case each select signal is forwarded to all segments above and each segment has an OR gate as wide as the number of segments that are located below.

35 If power consumption is not a concern, only the segment directly above the selected one will need to be disconnected from the segment below as indicated in Fig. 14C. Here the coupling signal is disabled for SELn-1.

DRAM cells with separate read and write word lines and clamped amplifier

The DRAM cell shown in Fig. 15 requires a larger chip area and more complex word line drivers, but offers advantages in two areas. The cell has the same advantages as the simpler cell described above in that the content of the cell is not disturbed during the refresh cycle. The fact that the word line voltage at reading has the full amplitude means that the read time is shorter than when the word line is left at an intermediate level. There is no feedback from the write transistor as the write word line is low during reading. If the refresh frequency is high enough, a stored ZERO will stay very close to Vss and a stored ONE will stay one threshold under WLW. Boosting the voltage of WLW can maintain the cell voltage of a stored ONE close to Vdd. This property makes the cell useful for control of logic, where the control node indicated in the figure connects to the logic to be controlled.

The amplifier can have all the forms as discussed before and can also be used in the segmented configuration. Bit lines need not be dedicated as indicated in the figure, but will in the non-dedicated case require dummy transistors on both bit lines under control of the address logic. The assumption is that the layout is more efficient with every other cell reversed around a vertical axle.

The amplifier in a different mode of operation can act as a latch, which is first reset, indicating a ZERO, and at reading a ONE is set. This requires that the cell with full cell voltage is strong enough to set the amplifier flip- flop. The segmented memory approach can still be taken.

Full Amplifier DRAM Configurations

In situations where fast access is of prime importance, the single inverting amplifier in each segment may be replaced with a full latch amplifier in each segment. This allows for simpler control with fewer control signals. The relative timing is however slightly more critical.

The DRAM with a full amplifier can operate in basically two modes. In Mode 1, shown in Fig. 16A, power is

always applied to the amplifier, and the bit lines are clamped to the same potential, which is the switching point of the latch. In the Mode 2, shown in Figs. 16B and 16C, power to the amplifier is turned off during the clamping and the first

5 portion of the read cycle. In Fig. 16B, the amplifier is OFF in standby mode with the bit lines clamped to a reference clamp level. In Fig. 16C, on the other hand, power is ON in standby leaving one bit line at Vdd and the other at Vss level. The single clamp transistor is turned on at about the same time as

10 power is turned off. If power is turned off before applying the clamp signal the bit lines will reach a voltage of approximately Vdd/2 due to charge sharing. If power is still on, the bit lines will eventually reach the switching point level of the amplifier latch. By designing the amplifier for a

15 switching point level of Vdd/2, the relative timing at this time is not critical. Power is turned on again, when the input signal has been integrated to a safe level, to guarantee that the amplification will go in the right direction.

The bit lines as shown in Figs. 16A-C are dedicated

20 to reading and writing, which means that all cells in a segment are oriented the same way, with the write transistors connected to the write bit line (BLW) and all read transistor connected to the read bit line (BLR). To allow more options for cell layout, it may be advantageous to have non-dedicated bit lines

25 as illustrated in Fig. 16D. There is a minor added cost in each amplifier in having two dummy transistor and controls to select the one corresponding to the address of the selected cell.

Figs. 17A-M show the result of simulations using a

30 0.5u NMOS process. The memory module consists of 8 segments per column, with 64 bits/segment, and 128 columns (65,536 bits).

Fig. 17A illustrates reading a ONE in a cell located in the top segment, with the read data propagating from BLR and

35 BLW in the top segment to BLR0 and BLW0 in the bottom segment. The data on BLR0 and BLW0 is forwarded to the I/O decoder at about 8 ns (not shown). Power in this case is initially turned off at the same time as clamping takes place (compare to Fig.

16C) and is turned on again when reading starts by bringing the word line (WL) to the first ramp level (1.8 V). The signal in the selected segment (the top segment in this case for worst case test) is amplified slightly before the COPL signal

5 connects the selected segment to the rest of the segments that were interconnected earlier by the signal COPLN. In this simulation the non-selected segments were clamped a little longer by the signal COPLN, which is 0.5 ns longer than COPL, the signal clamping the selected segment. The stored energy in

10 the bit lines of the selected segment is now transferred to the lower segments at the same time as each of the segments contributing to the amplification of the transmitted signal. At about 6 ns, the delayed signal has reached the bit lines BLR0 and BLW0 of the bottom segment and is reaching half of

15 full amplitude at a delay of less than 7.5 ns. All bit lines are very close to the supply rails at the nominal end of the cycle (10 ns). The voltage difference between BLW0 and BLR0 at 8 ns is more than 2.5 V, which is sufficient to supply an output signal via decoder and output amplifier within 2 ns.

20 The initial cell voltage was assumed to have degenerated due to leakage after the last refresh cycle to 2.5 V, but is refreshed in the shown read cycle to 3.5 V. The cell capacitance in the simulation is 2e-15 Farad plus the gate capacitance of the sense transistor (MS in Figures 1-3). There is a small

25 coupling from the word line, which can be noticed in the figure, so the final cell voltage when WL reaches Vdd is slightly below 3.5 V. The maximum voltage is determined by the max value of WL less the threshold voltage. The short channel process has a strong body effect on the threshold. The cell

30 current in the read phase increases very slowly with the cell voltage, when above 2.5 V, so boosting the WL voltage above Vdd would have a small effect on speed, but would of course allow for more leakage in the cell capacitor.

Fig. 17B shows the reading of a stored ZERO

35 degenerated due to leakage from 0 V to 1 V. Only a limited number of control signals are shown in this figure. It is a clear indication in this figure that the switching point of the amplifier is below the clamp level defined by the charge

sharing between the bit lines. This causes the bit lines to drift in the negative direction, before the positive current from the dummy transistor less the low cell current causes the read bit line (BLR) to go positive. In the full amplifier this
5   drift is acceptable as long as the device parameters in the amplifier are consistent.

In the write operation (Fig. 17C) the data to be written is applied via the column decoder to BLW0 and/or BLR0 on the bottom segment. This signal is large enough to be
10   amplified and propagate to the selected segment, which has started a normal read sequence, and override this signal. Fig. 17C shows how a stored ZERO is written over by a ONE. BLR0 and BLW0 are separated by the input signal already during the clamping operation. With all non-selected segments
15   interconnected, the signal is quickly propagated to the selected segment, causing the direction of BLR and BLW to reverse, so that BLW will drive the cell voltage to 3.5 V when WL goes high. Fig. 17D illustrates how a stored ONE is written over by a ZERO.
20   Turning off power during clamping saves power, as shown in Figs. 17E-N, but adds to circuit complexity. Figs. 17E-G all have a first ramp level of 1.8 V, while the following simulation used a first ramp level of 1.5 V. The circuit for the dummy current control adjusts the dummy current
25   to an optimal value for a given first ramp level. The current difference between the cell current and the dummy current varies very little depending on the ramp level, but there is less dependence on device parameter variations at the lower first ramp level value.
30   Comparing the initial clamp level of Fig. 17E with the previous simulations shows that the clamp level has dropped. This is caused by a change of the cell parameters equalizing the sizes of the read and write transistors and their junction areas. The higher junction capacitance at 0 V,
35   compared to 5 V, brings the clamping level below 2.5 V. The switching point should also be set equal to the clamp level. The clamp pulses CLNP and CLMPN in Fig. 17E terminated before the word line had reached the first ramp level. This could

cause the dummy current, if it had built up earlier to start the displacement of the read bit line prematurely. The margins otherwise are sufficient for correct reading of a ONE, as the figure shows. To eliminate the possibility of the problem

5 discussed above, the length of the clamp pulses in the following simulations were increased. Fig. 17F also displays the dummy voltage, that is applied to the gate of the P-channel dummy transistor. The dummy voltage is also close to its maximum negative value at the end of the clamp signals. A ONE

10 is read in Fig. 17F and a ZERO is read in Fig. 17G. Figs. 17H and 17I show normal read restore cycles for ZERO and ONE respectively.

It should be pointed out that in all cases a refresh cycle is the same as a read cycle except that all segments

15 operate simultaneously fully disconnected from each other. The signals COPL and COPLN are therefore inactive in the refresh cycles. All segments are enabled and the low order address bits addresses "the same" word line in each segment. With the timing for the clamping operation used in Figures 17E-Q, CLMP

20 and CLMPN have the same duration and is in reality the same pulse. Figs. 17J and 17K illustrate refresh cycles with the segments disconnected restoring ZEROs and ONEs. Comparing Figs. 17H and 17J shows how the bit lines move faster without the added load of the additional segments. This indicates that

25 the refresh cycles can be made shorter than a read or write cycle. Figs. 17L and 17M show the write cycles for storing ZERO and ONE.

Memory Refresh Interruption without Delay

30 One reasons why SRAMs are typically preferred over conventional DRAMs is that synchronous systems require that a memory request is served without the delays that a DRAM refresh cycle would cause. A great advantage to the DRAM of the present invention is that an ongoing refresh cycle can be

35 interrupted at any time during the cycle without destroying the stored data. Data stored in the cell capacitor remains intact most of the time or is refreshed during the end of the cycle. An ongoing refresh cycle can be stopped as long as the word

line is brought low as the first step. Figs. 18A-B show the simulations of interrupting refresh cycles after 6 ns and starting read cycles on a different address, while Figs. 18C-D show an interrupting write cycle.

5   In these simulations the CLMP signal starts 1 ns later than in the previous simulations, but ends at the same time. The shorter pulse is sufficient for the clamping of the bit lines and the delay makes sure that the bit line voltages have not changed before the word line active during the refresh

10 has been brought down. It is assumed that the refresh and the reading or writing is done in the top segment, with the bit lines BLR and BLW. The word line addressed during the refresh is WLA and the word line addressed during read or write is WLB. Data input and output connects to the bit lines of the bottom

15 segment BLR0 and BLW0. The interrupting memory cycle starts with signal INT, which in the simulation was used to force the WLA low, while the decoder was selecting WLB. In Fig. 18A the cell addressed by WLA (CELLA) is in the ZERO state, starting at 0.5 V and during the read phase of the refresh cycle is lifted

20 to 0.6 V and returned to 0.5 V when WLA is turned off (capacitive coupling between WLA and the cell capacitor). The cell to be read is at 2.5 V and is brought to 3.5 V at the end of the interrupting memory cycle. In Fig. 18B the cell to be refreshed (CELLA) is at 2.5 V and the cell to be read in the

25 interrupting cycle (CELLB) is at 0.5 V. The reading returns CELLB to 0 V.

       In Figs. 18C-D the refresh cycles are interrupted by write cycles. In Fig. 18C the cell to be refreshed (CELLA) is at 2.5 V and remains at the same level after interrupting the

30 refresh before WLA has been brought to 5 V. CELLB starts at its previous value of 0.5 V and at the end of the write cycle is at 3.5 V. In Fig. 18D CELLA remains at the 0.5 V it was at the start of the refresh cycle, while the voltage of CELLB is changed from its previously stored value of 3.5 V to 0 V. In

35 the write operation a signal COPLIN occurring at the same time as the signal COPLN is applying the input data to the bottom segments bit lines BLR0 and BLW0. As can be seen these bit lines proceed in time the bit lines BLR and BLW in the top

segment, contrary to the situation in Figs. 18A-B, where BLR and BLW come first.

Figs. 18A-D illustrate how a memory cycle can interrupt a refresh cycle without destroying the stored data for an arbitrarily chosen time in the refresh cycle. In the example a previously degenerated cell voltage was chosen (some time after last refresh). An incomplete refresh due to an interrupting memory access will cause further degeneration until new uninterrupted refresh cycle is performed. The

10 address of the cells of the interrupted refresh would be retained during the memory access (accesses) and would be the first address used when the refresh cycle is resumed. With the very low refresh frequency compared to the high memory access rate very large blocks of data with very short interruptions

15 would have to be transferred before the refresh integrity would be in danger. As a backup for all eventualities a long delay circuit could be used to interrupt the flow occasionally.

Related to the idea of interrupting an ongoing refresh cycle at any time without destroying the stored data is

20 the concept of performing a fast burst read. A fast burst read operation involves reading data from the memory cell without completing a refresh cycle following the read operation. In the DRAM of the present invention, this is possible since the data stored in a memory cell is not destroyed when the contents

25 of the cell are read, eliminating the need to refresh the contents of the cell. By reading data from a memory cell without a subsequent refresh cycle, the amount of time needed to perform a read operation is typically halved. Thus, the possible frequency of read cycles is doubled, improving the

30 efficiency of memory operation. Of course, it should be understood that a standard DRAM refresh cycle must still be performed. The fast burst read operation essentially turns the customary read/refresh operation into a strict read operation, allowing a read operation to be performed without an associated

35 refresh operation.

DRAM Cluster Cell

In another embodiment, the DRAM of the present invention may include more than one storage capacitor associated with a single pair of read and write transistors. Fig. 19A shows a simplified circuit diagram 1900 of a "cluster

5   cell" using four storage nodes C1-C4. Instead of having one storage node at the common (COM) node, any of the four node capacitors C1-C4 can be connected to COM during a read or write operation. The COM node can be looked upon as a very short bit line in a conventional DRAM and all the techniques for layout

10   and processing can be applied to the cells on the COM "bit line."

The number of cells that can be placed in the cluster depends on the relation between the memory cell capacitance and the capacitance of the COM node. The diagram in Fig. 19A shows

15   four memory cells in the cluster, merely by way of example. The COM node capacitance includes the junction capacitance of the COM bit line, the source junction of write transistor 1902, the drain junction on CLMPCOM transistor 1904 and the gate of sense transistor 1906. The COM bit line may have relatively

20   high resistance, so minimum active junctions can be used.

In operation, the primary DRAM cell, which normally reads and restores the charge stored in a capacitor connected to the COM node, has been modified so that any of the cell capacitors C1-C4 on the COM bit line can be connected to COM

25   before the reading and restoring of the charge on the selected cell. Due to charge sharing between a cell capacitor and the COM node, there is a slight modification of the cell voltage after the connection. The actual voltage on the COM node depends on whether a ONE or a ZERO was read in the previous

30   cycle. In order to make the effect of the charge sharing more predictable, the COM node is clamped to a fixed level between cycles. Once the cell is connected to COM, a normal cycle is then executed. Fig. 19A also shows a portion of a word line decoder 1910 coupled to master word line WL and four slave word

35   lines WLC1, WLC2, WLC3 and WLC4. A number of global signals are also shown that define active time for the different horizontal lines. An active subcell is selected by a decoder (not shown) and applied at a specified time on C1TIMEB,

C2TIMEB, C3TIMEB or C4TIMEB. An amplifier 1920 is also shown in Fig. 19A, and its performance is similar to that described above.

Fig. 19B shows the timing relationship between the
5 different signals. For example, the reading and restoring of a ZERO and of a ONE is shown. The charge sharing effect between the cell capacitor and the COM node, when the COM node capacitance when the capacitances are equal is also illustrated.

10 With the COM node clamped at 2 V between cycles, the displacement on the stored value while reading is the same, but in opposite directions. The cell current to BLR is normally limited by the low WL voltage on read transistor 1908 and therefore increases less than linearly with the cell voltage.
15 The cost of adding a separate write word line 1920 is minimal to address this limitation, as shown in an alternate cluster cell configuration 1950 in Fig. 19C. The high WLR voltage (5 V) renders the cell current quadratically dependent of the active voltage on sense transistor 1906 (the COM node). The
20 much higher read current on BLR makes for an even faster read operation and higher margins. The timing diagrams for the circuit of Fig. 19C is shown in Fig. 19D.

If a cluster cell is read before a refresh cycle, the cell voltage will have been modified by charge sharing with the
25 COM node, as described above. A second reading may still be correct if the COM node capacitance is very small compared to the cell capacitor. The modification is however progressive from cycle to cycle, so a burst read could only be permitted once or twice before the cell is refreshed. The same
30 restriction applies to interrupted refresh cycles for reading the same or other addresses and writing to other addresses. Note that the two tier configuration described earlier is still valid in the cluster cell as long as the second tier design takes into account the charge sharing effect.

35

The above description is illustrative and not restrictive. Many variations of the invention will become apparent to those of skill in the art upon review of this disclosure. The scope of the invention should, therefore, be

5   determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

WHAT IS CLAIMED IS:

1   1.   A dynamic random access memory comprising:
2   a plurality of memory cells;
3   a plurality of word lines coupled to said plurality
4   of memory cells; and
5   a plurality of bit lines coupled to said plurality of
6   word lines, wherein each bit line is divided into a plurality
7   of bit line segments separated by pass transistors, each bit
8   line segment coupled to at least one of said plurality of
9   memory cells and an associated sense amplifier.

1   2.   The dynamic random access memory of claim 1
2   further comprising:
3   a plurality of segment decoders coupled to said
4   plurality of bit line segments;
5   a plurality of word line decoders coupled to said
6   plurality of word lines; and
7   a plurality of address lines coupled to said
8   plurality of word line decoders and said plurality of segment
9   decoders.

1   3.   The dynamic random access memory of claim 1
2   wherein said associated sense amplifier further comprises a
3   latching amplifier.

1   4.   The dynamic random access memory of claim 1
2   wherein said associated sense amplifier further comprises an
3   amplifier including at least one inverter.

1   5.   The dynamic random access memory of claim 4
2   wherein a first bit line segment is coupled to an adjacent bit
3   line segment such that said inverter in said first bit line
4   segment and said inverter in said adjacent bit line segment
5   correspond to form a full amplifier.

1   6.   The dynamic random access memory of claim 1
2 further comprising a clamp transistor coupled to a bit line and
3 a clamp line, said clamp line having a clamp voltage level.

1   7.   The dynamic random access memory of claim 1
2 wherein one of said plurality of bit line segments further
3 comprises:
4   a storage capacitor for storing charge representative
5 of a stored value;
6   a first, write transistor coupled to a first plate of
7 said storage capacitor and a first bit line;
8   a second, read transistor having a first terminal
9 coupled to a first terminal of a third transistor and a second
10 terminal coupled to a second bit line, said third transistor
11 comprising a gate coupled to said first plate of said storage
12 capacitor, a first one of said plurality of word lines coupled
13 to a gate of said write transistor;
14   a fourth transistor coupled to said first bit line
15 and a clamp line, said clamp line having a clamp level
16 intermediate to ground and a switching threshold of said write
17 transistor; and
18   a sense amplifier coupled to said first and second
19 bit lines.

1   8.   The dynamic random access memory of claim 7
2 wherein a second one of said plurality of word lines is coupled
3 to a gate of said read transistor.

1   9.   The dynamic random access memory of claim 7
2 wherein said sense amplifier further comprises:
3   a first transistor coupled to said first and second
4 bit lines and a first potential;
5   a second transistor coupled to said first transistor,
6 said first bit line and a second potential;
7   a third transistor coupled to said first and second
8 bit lines and said first potential; and
9   a fourth transistor coupled to said first and second
10 bit lines and said second potential.

1   10. The dynamic random access memory of claim 7
2 wherein said sense amplifier further comprises:
3   a first transistor coupled to said first and second
4 bit lines and a first potential; and
5   a second transistor coupled to said first transistor,
6 said first bit line and a second potential.

1   11. The dynamic random access memory of claim 1
2 further comprising shared bit lines.

1   12. A programmable logic device including the
2 dynamic random access memory of claim 1 for control of said
3 programmable logic device.

4   13. The programmable logic device of claim 12
5 further comprising:
6   a pass transistor coupled to the dynamic random
7 access memory; and
8   a storage capacitor formed under said pass
9 transistor.

10   14. The programmable logic device of claim 12
11 further comprising:
12   a pass transistor coupled to the dynamic random
13 access memory; and
14   a storage capacitor formed above said pass
15 transistor.

1   15. A two-tier dynamic random access memory cell
2 comprising:
3   a first tier memory with a first addressing circuit;
4 and
5   a second tier memory with a second addressing circuit
6 for reading contents of said first tier memory, wherein said
7 first tier memory receives data input simultaneously with the
8 reading of contents of said first tier memory by said second
9 tier memory.

1   16. The two-tier dynamic random access memory of

2 claim 15 further comprising:

3   a storage capacitor for storing charge representative

4 of a stored value;

5   a first, write transistor coupled to a first plate of

6 said storage capacitor and a first bit line;

7   a second, read transistor having a first terminal

8 coupled to a first terminal of a third transistor and a second

9 terminal coupled to a second bit line, said third transistor

10 comprising a gate coupled to said first plate of said storage

11 capacitor, a gate of both said read transistor and said write

12 transistor coupled to a word line;

13   a fourth transistor coupled to said first plate of

14 said storage capacitor;

15   a fifth transistor having a first terminal coupled to

16 said four transistor and a second terminal coupled to a third

17 bit line;

18   a first sense amplifier coupled to said first and

19 second bit lines; and

20   a second sense amplifier coupled to said third bit

21 line.

1   17. A dynamic random access memory comprising:

2   a storage capacitor for storing charge representative

3 of a stored value;

4   a first, write transistor coupled to a first plate of

5 said storage capacitor and a first bit line;

6   a second, read transistor having a first terminal

7 coupled to a first terminal of a third transistor and a second

8 terminal coupled to a second bit line, said third transistor

9 comprising a gate coupled to said first plate of said storage

10 capacitor, a gate of both said read transistor and said write

11 transistor coupled to a word line;

12   a fourth transistor coupled to said first bit line

13 and a clamp line, said clamp line having a clamp level

14 intermediate to ground and a switching threshold of said write

15 transistor; and

16   a sense amplifier coupled to said first and second
17 bit lines,
18   wherein a refresh cycle of the dynamic random access
19 memory may be interrupted with a memory read cycle without
20 destroying data stored in the dynamic random access memory, and
21   wherein a refresh cycle of the dynamic random access
22 memory may be interrupted with a memory write cycle without
23 destroying data stored in non-addressed columns of the dynamic
24 random access memory.

1   18. The dynamic random access memory of claim 17
2 wherein a fast burst read cycle is performed without a
3 subsequent associated memory refresh operation.

1   19. A method of forming a capacitor on a
2 semiconductor support wafer comprising the steps of:
3   etching a recessed area in the support wafer;
4   implanting a dopant layer in the support wafer;
5   growing a thin oxide as a dielectric layer of the
6 capacitor;
7   depositing a polysilicon layer until said recessed
8 area is filled;
9   polishing the support wafer;
10   growing an epitaxial layer on the support wafer;
11   implanting oxygen to a depth equal to said epitaxial
12 layer; and
13   forming an insulating layer of silicon dioxide.

1   20. The method of claim 19 wherein the step of
2 etching a recessed area further comprises the step of etching a
3 P-type recessed area in the support wafer.

1   21. The method of claim 19 further comprising the
2 step of heat-treating the support wafer.

1   22. The method of claim 19 wherein said step of
2 implanting a dopant layer further comprises the step of
3 implanting an N+ layer in the support wafer.

23. A method of submerging a transistor in a semiconductor support wafer comprising the steps of:

etching a recessed area in the support wafer;

depositing a polysilicon layer in the support wafer;

polishing the support wafer;

implanting a dopant layer in the support wafer;

growing an epitaxial layer on the support wafer; and

implanting a gas in the support wafer.

24. The method of claim 23 further comprising the step of heat-treating the support wafer.

25. The method of claim 23 wherein said step of implanting a gas further comprises the step of implanting oxygen in the support wafer.

26. The method of claim 23 wherein said step of implanting a gas further comprises the step of implanting nitrogen in the support wafer.

27. A method of forming a capacitor on a semiconductor support wafer comprising the steps of:

forming a diffusion area in the support wafer;

implanting a first gas in the support wafer to form a dielectric layer;

etching a trench in the support wafer to isolate the capacitor;

growing an epitaxial layer; and

implanting a second gas to separate said epitaxial layer from the support wafer.

28. The method of claim 27 further comprising the step of heat-treating the support wafer.

29. The method of claim 27 wherein said step of implanting a first gas in the support wafer further comprises the step of implanting oxygen to form said dielectric layer.

1   30. The method of claim 27 wherein said step of
2 implanting a first gas in the support wafer further comprises
3 the step of implanting nitrogen to form said dielectric layer.

1   31. A method of forming a submerged transistor and a
2 capacitor on a semiconductor support wafer comprising the steps
3 of:
4   forming a first and second diffusion area in the
5 support wafer;
6   implanting a first gas in the support wafer to form a
7 dielectric layer coupling said first and second diffusion
8 areas;
9   etching an isolation trench in the support wafer;
10   implanting a dopant layer to form the capacitor;
11   growing an epitaxial layer; and
12   implanting a second gas to separate said epitaxial
13 layer from the support wafer.

1   32. The method of claim 31 further comprising the
2 step of heat-treating the support wafer.

1   33. The method of claim 31 wherein said step of
2 implanting a first gas in the support wafer further comprises
3 the step of implanting oxygen to form said dielectric layer.

1   34. The method of claim 31 wherein said step of
2 implanting a first gas in the support wafer further comprises
3 the step of implanting nitrogen to form said dielectric layer.

1   35. The method of claim 31 wherein said step of
2 implanting a dopant layer further comprises the step of
3 implanting an N+ layer.

1   36. A method of operating a dynamic random access
2 memory comprising a plurality of memory cells, a plurality of
3 word lines coupled to said plurality of memory cells and a
4 plurality of bit lines divided into bit line segments, each
5 segment coupled to at least one of the plurality of memory

6   cells and an associated sense amplifier, the method comprising

7   the steps of:

8   storing a charge representing a data value in a

9   memory cell capacitance;

10   activating a word line to operatively couple said

11   memory cell to a selected bit line segment, said charge in said

12   memory cell either permitting or not permitting a selected

13   current to flow to said selected bit line segment depending

14   upon said data value;

15   sensing a difference in voltage between two sense

16   nodes with the sense amplifier, said voltage difference

17   established by said selected current flow and a reference

18   current flow to a sense node; and

19   activating said word line to refresh said charge in

20   said memory cell capacitance in said selected bit line segment.

1   37. The method of claim 36 further comprising the

2   step of reading said charge from said memory cell by

3   propagating a signal representing said voltage difference from

4   said selected bit line segment to remaining bit line segments

5   in a bit line.

1   38. A dynamic random access memory comprising:

2   a plurality of memory cells at the intersections of

3   word lines and bit lines, each memory cell comprising:

4   a plurality of storage capacitors for storing

5   charge representative of a stored value;

6   a plurality of switch transitors, each coupled

7   to a first plate of one of said plurality of storage

8   capacitors;

9   a write transistor coupled to said plurality of

10   switch transistors at a common node and a first bit

11   line;

12   a sense transistor having a gate coupled to said

13   common node and a source coupled to a second plate of

14   each of said plurality of storage capacitors; and

15   a read transistor coupled to said sense

16   transistor and a second bit line; and

17   a sense amplifier coupled to said first and second
18 bit lines.

1   39.   The dynamic random access memory of claim 38
2 further comprising:
3   a clamp transistor coupled to said common node; and
4   a clamp voltage line coupled to said clamp
5 transistor, said clamp transistor coupling a common clamp
6 voltage level on said clamp voltage line to said common node
7 prior to a read or write operation of said memory cell.

1   40.   The dynamic random access memory of claim 38
2 further comprising:
3   a first word line coupled to a gate of said write
4 transistor; and
5   a second word line coupled to a gate of said read
6 transistor.

1   41.   The dynamic random access memory of claim 38
2 wherein said each of said bit lines is divided into a plurality
3 of bit line segments separated by pass transistors, each bit
4 line segment coupled to at least one of said plurality of
5 memory cells and an associated sense amplifier.

1   42.   A dynamic random access memory comprising:
2   a plurality of memory cells at the intersections of
3 word lines and bit lines, each memory cell comprising:
4   a storage capacitor for storing charge
5   representative of a stored value;
6   a write transistor coupled to a first plate of
7   said storage capacitor and a first bit line;
8   a sense transistor having a gate coupled to said
9   write transistor and said first plate of said storage
10   capacitor;
11   a read transistor having a first terminal
12   coupled to said sense transistor and a second
13   terminal coupled to a second bit line;

14   a first word line coupled to a gate of said
15   write transistor; and
16   a second word line coupled to a gate of said
17   read transistor; and
18   a sense amplifier coupled to said first and second
19 bit lines.

43. **A dynamic random access memory, substantially** as herein described with reference to, or as shown in, the accompanying drawings.

44. A method of operating a dynamic random access memory, substantially as herein described with reference to the accompanying drawings.

45. A method of forming an integrated circuit component, substantially as herein described with reference to, or as illustrated in, any of the accompanying drawings.

Application No: GB 9704265.9   **Examiner:** Brian Ede
Claims searched: 1-14   **Date of search:** 15 May 1997

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

     UK Cl (Ed.O): G4C(C700L, C706, C11409A, C11409L)

     Int Cl (Ed.6): G11C 7/00 11/409

Other: Online: WPI

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
X	EP 0431865 A2 (NEC CORP) see Figure 5	1-4 at least
X	EP 0282650 A1 (NEC CORP) see Figure 4	1-4 at least
X	EP 0239225 A2 (MITSUBISHI) see Figure 5	1 and 4 at least
X	WO91/01557A2 (SAMSUNG) see 14, 20, 22, 24, 26, 36 Figure 1	1-4 at least
X	US 5016224 (TOSHIBA) see Figure 1	1-4 at least
X	US 4920517 (MATSUSHITA) see Figure 6	1 and 2 at least

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.